Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Partial and 100% Reinforcement Under Chance and Skill Conditions: W. H. JAMES AND J. B. ROTTER 397
UCS Intensity and the Associative (Habit) Strength of the Eyelid CR: K. W. SPENCE, D. F. HAGGARD, AND L. E. ROSS 404
A Further Test of the Hypothesis of Autonomous Memory Trace Change: S. H. LOVIBOND .. 412
Effects of Window Size and Two Conditions of Judgment on the Reproduction of Objective Velocity: O. W. SMITH 416
The Dependence of Probability of Response on Size of Step Interval in the Method of Limits: J. BRACKMANN AND G. COLLIER 423
Nonindependence of Successive Responses at the Visual Threshold as a Function of Interpolated Stimuli: G. COLLIER AND W. S. VERPLANCK 429
Dimensional Analysis of Movement Reactions: E. A. FLEISHMAN 438
The Dimensional Analysis of a New Series of Facial Expressions: T. ENGEN, N. LEVY, AND H. SCHLOSBERG 454
Task-Influence and the Stability of Generalized Expectancies: J. F. RYCHLAK 459
Ease of Attainment of Concepts as a Function of Response Dominance Variance: J. L. FREEDMAN AND S. A. MEDNICK 463
Isolation Effects: Stimulus and Response Generalization as Explanatory Concepts: S. E. NEWMAN AND E. SALTZ 467
Effects of Strength of Drive on Learning and on Extinction: H. BARRY, III 473
Half-Loudness Judgments Without Prior Stimulus Context: W. R. GARNER 482
Age Changes and Information Loss in Performance of a Pursuit Tracking Task Involving Interrupted Preview: S. GRIEW 486
Effects of Previously Associated Annoying Stimuli (Auditory) on Visual Recognition Thresholds: J. HOCHBERG AND V. BROOKS 490
Position Preference and Discrimination Learning: M. H. GOER 492
The Effects of Degree of Learning and Problem Difficulty on Perseveration: R. THOMPSON ... 496
Conditioning of the Human Pupillary Dilation Response as a Function of the CS-UCS Interval: A. A. GERALL AND J. K. WOODWARD 501

American Psychological Association

Vol. 55, No. 5 May 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor: Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices in this Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs supplementary-material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 55, NO. 5 MAY, 1958

PARTIAL AND 100% REINFORCEMENT UNDER CHANCE AND SKILL CONDITIONS[1]

WILLIAM H. JAMES[2] AND JULIAN B. ROTTER

Ohio State University

Previous experiments on partial and 100% reinforcement have usually indicated that training under partially reinforced conditions produces greater resistance to extinction than 100% conditions. Most of the work in this area, however, has been conducted with animal *Ss* and has dealt with relatively simple responses where possible symbolic behavior was held to a minimum. Studies with human *Ss* likewise have tended to use simple conditioning procedures where symbolic behavior and alternative responses were minimized. An analysis of the methodologies used in this research indicates that the learning situation is one where *S* is likely to perceive reinforcements as being beyond his control and primarily contingent upon external conditions. In learning situations where *S* can perceive reinforcements as being within his control or due to some stable skill which he can utilize, the effects of reinforcement on both acquisition and extinction might follow different functions from those in the traditional situation.

Many of the complex behavioral problems with which the clinical psychologist must deal involve learning with *Ss* who are moving toward long-range goals and who can perceive themselves as having some control over the pathways and potential reinforcements relating to these goals.

Thus, the learning curves of a college student seeking reinforcements in the form of good grades, who perceives these reinforcements as dependent on skills which he has or can develop, might be quite different from those obtained when the student is in a laboratory situation and reinforcement is in the form of an air blast to establish a conditioned eyelid response.

It is generally acknowledged that application of learning theory to human behavior will require the postulation of additional variables, especially those involving the use of symbolic behavior as a means by which situations are categorized. It was postulated that a possibly important dimension by which situations are categorized is the extent to which the reinforcements are controlled by external or chance factors versus the extent to which they can be

[1] Read, in part, at the Midwestern Psychological Association meetings, Spring, 1957.

[2] Now at the University of Wisconsin.

related to internally controlled or skill factors.

Some evidence of the importance of this variable has been provided by recent studies designed in the framework of social learning theory (**5**). Phares (**4**), utilizing operations for measuring expectancy, gave *Ss* instructions which emphasized that success on a task was due to skill. Other groups were given instructions which emphasized that success on the same task was due to chance. He found significantly greater changes in expectancies in his skill groups, although all groups had the same number and pattern of reinforcements. His skill groups were also characterized by significantly fewer unusual shifts in expectancies. (An unusual shift is defined as a raise in expectancy following negative reinforcement or a decrease in expectancy following positive reinforcement.) In another study by Neff (**3**) suggestive evidence was obtained that the extinction of expectancies under skill conditions would follow a different function from those found in the usual experimental conditions. Using a task which emphasized that success was due to *S*'s skill, Neff found no significant differences in trials to extinction between a 100% and a 50% random reinforcement group. In social learning theory, expectancy is defined as "the probability held by the individual that a particular reinforcement will occur . . . in a specific situation or situations" (**5**, p. 107).

A recent study by Wyckoff and Sidowski (**6**) bears some relation to this problem. They obtained results for a situation involving intermittent reinforcement somewhat different from those usually found. In their experiment *S* had a motor tracking task to accomplish involving anticipation of the direction of an E-controlled moving target. They attributed their results to a reduced tendency to respond to the task as a "guessing" problem as the experiment progressed.

The present study is a further attempt to investigate differences in the development and extinction of expectancies under conditions of external control (or chance) and internal control (or skill). On the basis of social learning theory (**5**) it was reasoned that, under chance conditions of 100% reinforcement, *S* received no cues of nonreinforcement during the training series. When extinction begins he utilizes the first nonreinforcements as cues that the situation has changed. Since control of the situation is perceived as external there is a sudden shift in expectancy and extinction is rapid. With partial reinforcement under chance conditions *S* receives cues of nonreinforcement as well as reinforcement. When extinction is begun nonreinforcement does not produce a new cue which he can use to discriminate a change in situations. Consequently, extinction is more gradual and more trials are required before the situation is recategorized. Another factor which also increases resistance to extinction under these conditions is *S*'s tendency to count and verbalize relationships. Thus, during acquisition he is likely to develop the hypothesis that a series of nonreinforced trials will increase the probability that a reinforcement is forthcoming. Consequently, during extinction he may raise his expectancy after a series of nonreinforced trials. With continued extinction several raises may occur which early in extinction may correspond to the original patterning of reinforcements during training. This was demonstrated in a study by Lasko (**2**).

It was reasoned that, under skill conditions, *S* would perceive the reinforcements and cues in the situation as being produced by factors controlled by himself. With 100% reinforcement followed by extinction he would be less likely to recategorize the situation as having changed, and sudden decreases in expectancy would not be obtained. With partial reinforcement under skill conditions the kind of "gambler's fallacy" phenomenon observed under chance conditions would not be obtained since *S* will attribute nonreinforcement to a lack of skill on his part rather than to luck factors and hypothesized

probabilities. Therefore, since the 100% group would be more resistant to extinction and the partial group less resistant, the differences usually found in chance-structured situations would not be as great. Thus, it was hypothesized that the superiority of a partially reinforced group in resistance to extinction would be significantly less under conditions of internal control of reinforcement (skill) than under conditions of external control of reinforcement (chance).

Although there was some evidence of this from the previously mentioned studies of Phares and Neff, no specific experiment to test for differences in trials to extinction under the chance and skill situations on the same task had been conducted. The purpose of the present study was to employ the same task for all groups of *Ss* and to vary only the instructions which were designed to emphasize that success on the task was due either to chance factors or skill factors.

Method

Subjects.—The *Ss* were 80 (23 males, 57 females) college freshmen students taking courses in introductory psychology. Subjects were randomly assigned to each of the four experimental groups making a total of 20 *Ss* in each group.

Task.—It was necessary to use a task which could be perceived as chance or skill depending upon the instructions. Pre-experimental testing indicated that, with detailed instructions, it was possible to use a simple card-guessing game. Varied instructions were tested and *Ss* were interviewed regarding their reactions to them and to the task. It was found that the majority of *Ss* responded differentially to chance or skill instructions, although there were wide individual differences. The experiment was then conducted with a new sample of *Ss* as described above.

Apparatus.—A manually operated tachistoscopic-like apparatus was used for presenting the cards to be guessed. A large number of 2.5 × 3-in. cards were marked with either an **X** or an **O**. These were arranged in sets of seven cards each and were in view of *S* when he entered the room. He was told that the cards would be placed in the apparatus and would be presented one at a time in sets of seven. It was emphasized that there were not necessarily equal numbers of **X**'s and **O**'s in any one set of cards. The cards which *S* saw were not really placed in the apparatus. In actuality a single card with a double face was used by *E*. By sliding this card a few inches in either direction it was possible to present either answer to *S* and control the reinforcement which he received. From *S*'s viewpoint each card of the ones he had originally seen was dropping in front of a closed shutter and he was to guess whether it was an **X** or an **O**. After he stated his guess the shutter opened and revealed the card. The shutter then closed and another card dropped into place. From *E*'s position there was a double-faced card with either **X** or **O** on it. As soon as *S*'s guess was stated, *E* slid the card in the desired direction and then opened the shutter. The *Ss* were seated 10 ft. from the apparatus.

Instructions.—"This is an experiment to see how well people can do at guessing a series of things when they have no directly observable cues to guide them." (Skill) "There is evidence that some people are considerably skilled at this." (Chance) "Scientists have shown so far that this is entirely a matter of luck."

"In this experiment we are using cards such as these which are marked with either an **X** or an **O**. A large series of these cards is placed in this apparatus. The cards are divided into sets of seven. There are not necessarily equal numbers of **X**'s and **O**'s in any set. Each time, a single card will slide in front of this closed shutter. You are to guess whether it is an **X** or an **O**. After you make your guess this shutter will open and you will know whether you were right or wrong. I will also be keeping score and will let you know how well you did at the end of each trial; that is, at the end of each set of seven cards.

"Just by blind guessing alone we would expect you to get about half the cards right on the average, but in order to score you must do better than chance. In order to receive a score you must get at least five or more cards correct out of seven. Any number of cards correct from five to seven will give you the same score. Any number of cards correct below five will earn no score. You must do better than chance in order to be successful.

"Before each trial I would also like you to estimate (skill) how well you feel you will do; (chance) how lucky you feel you will be. You can estimate this on a scale going from 0 to 10. For example if you feel fairly sure that you will

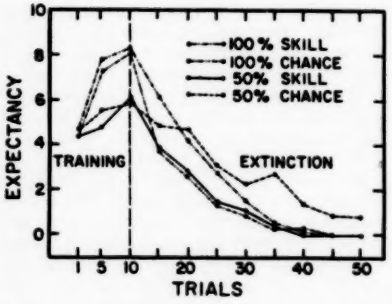

FIG. 1. Acquisition and extinction for 100% and 50% chance and skill groups ($N = 20$ per group).

(do well) (be lucky); that is, that you will do better than chance and get five or more cards correct, you might rate yourself with a 9 or 10. If you feel moderately sure that you will (succeed) (be lucky) you might rate yourself with a 5 or 6. If you feel pretty sure you (won't be successful) (won't be lucky) you might rate yourself with a 0 or 1. Use any of the numbers on the scale from 0 to 10 to indicate how you feel you will do. Now these estimates that you make before each trial can also affect your total score. If you are successful on the trial the estimate which you made will be added to your total score. If you are not successful the estimate will be subtracted from your total. Thus, it is important that you select your estimates carefully on the 0 to 10 scale and that they correspond closely with how well you feel you will do. The idea, of course, is to get as high a total score as you can. At the end of the experiment I will let you know how you did in comparison to other people who have taken this experiment."

Special skill group instructions.—"Although the cards are not arranged in any fixed order and there is no way to see a particular card beforehand, we have found that some people are considerably skilled at this. They can do consistently better than chance. Do you have any questions so far? All right—do as well as you can and let us see if you are skilled at this sort of thing. Now, before we begin would you make an estimate on the scale from 0 to 10 of the degree to which you feel you will succeed on this first trial."

Special chance group instructions.—"The cards are not arranged in any fixed order and there is no way to see a particular card beforehand. We have found that successful guessing is purely a matter of luck. Although at times people can do well, the results in the long run are like those obtained in a coin-tossing situation. Do you have any questions so far? All right—do as well as you can and let us see if you can get a high score on this.

"Now before we begin would you make an estimate on the scale from 0 to 10 of how lucky you feel you will be on this first trial."

Procedure.—Subjects in Group A, the 100% chance group, were then given 10 reinforced trials on which they received either 5 out of 7 or 6 out of 7 correct guesses on a trial.[3] After Trial 10, the extinction series was begun and *S*s were given either 3, 2, or 1 reinforcements out of 7 within each trial. After extinction Trial 25, only 1 or 2 reinforcements per trial were given. Before each trial, *S* was asked to state his expectancy for success on the 11-point scale. He was considered extinguished when his verbalized expectancy fell to either 0 or 1 on three consecutive trials. Group B, the 50% chance group, was given 10 acquisition trials in which half the trials were reinforced in a random pattern. After Trial 10, *S*s received the same extinction series as did Group A.

Group C, the 100% skill group, experienced the same experimental conditions as did Group A except that C was given the skill instructions while A was given the chance instructions. Group D, the 50% skill group, experienced the same experimental conditions as did B except that D was given skill instructions while B was given chance instructions.

RESULTS

Acquisition and extinction curves for the four groups are shown in Fig. 1. There were no significant differences in initial expectancies among the groups. All groups began in the 4.5 expectancy range. The acquisition curves of the 100% groups are fairly similar, as are the 50% groups, and differences between similar groups at Trial 10 are not significant.

[3] It should be remembered that *S*s were clearly instructed that a score of 5, 6, or 7 on a trial was a correct trial and all were scored equivalently as one correct trial. The 100% group were reinforced as doing better than chance on every training trial.

TABLE 1

TRIALS TO EXTINCTION

Group	Trials	*SD*
A (100% chance)	15.55	9.86
B (50% chance)	29.05	9.41
C (100% skill)	22.90	4.84
D (50% skill)	19.75	7.27

The mean number of trials to extinction and *SD*'s for each group can be seen in Table 1. As can be noted from the large *SD*'s, especially in the chance groups, there were wide individual differences among *Ss*. This can be attributed to the possibility that the instructions were not as effective for some *Ss* as they were for others and that some *Ss* operated according to their own hypotheses rather than the instructions. The data were tested with *t* tests for unequal variances as described in Dixon and Massey (**1**) since Bartlett's test for heterogeneity of variance indicated significant differences in group variances.

Results of *t* tests for trials to extinction between groups are shown in Table 2. Under chance conditions of partial reinforcement, Group B[4] required almost twice the mean number of trials to extinction as did Group A (100%). There were no significant differences between the variances of these groups, and the *t* value was significant beyond the .01 level.

[4] It is interesting to note that in Group B there were five *Ss* who did not extinguish after 40 extinction trials. At extinction Trial 40 these five *Ss* still stated a mean expectancy of 3.2. Since 40 were selected as the cut-off point for extinction because of time limitations these *Ss* were attributed with 40 trials. This is the only group in which *Ss* did not extinguish. In Fig. 1 the curves beyond Trial 40 are interpolated.

Under skilled conditions with Groups C and D a reversal took place. The 100% group was slightly superior to the 50% group in resistance to extinction. This difference was not significant, however. The variances between these two groups were significantly different with the 50% group showing the larger variance.

Comparisons made between Groups A and B revealed that the 100% skill group took significantly more trials to extinguish than did the 100% chance group. Despite unequal variances between these groups, the *t* test was significant. With the 50% groups, B and D, the reverse effect took place. Under skill conditions the 50% group took significantly fewer trials to extinguish than did the 50% chance group. There were no significant differences in variances between these groups. An analysis of individuals' scores in the 50% chance groups reveals that a number of *Ss* in this group had increments in expectancies at varying points during the extinction series. These increments did not occur at the same time for different *Ss* and consequently the curve based on mean expectancies does not indicate the alternate falling and rising effects. This phenomenon did operate to produce leveling effects in the group curve.

TABLE 2

DIFFERENCES BETWEEN GROUPS

Groups	Diff.	*t*	*P*
A − B	13.50	4.43	<.001
C − D	3.15	1.61	>.10
A − C	7.35	2.99	<.01
B − D	9.30	3.50	<.001
(A − B) − (C − D)	16.65	4.63	<.001

Discussion

Although these results are based on small samples and there were wide individual differences in each group (but considerably lower in the 100% skill group) these findings support the hypothesis of different extinction times and rates under chance and skill conditions. The wide individual differences in extinction were thought to be a function of differences in response to the instructions. There is some evidence in the study of Phares (4) that individuals vary in the extent to which they generally tend to perceive reinforcements as being internally or externally controlled. It is therefore possible that some individuals having strong general trends to see situations either in terms of being chance determined or skill determined would act on these general characteristics in a manner to offset the effect of the instructions. That is, there were conceivably some *Ss* in the chance groups who responded to the task as a skill situation despite instructions, while some *Ss* in the skill groups responded to it as being chance determined.

The most important implication of the study is that theorists attempting to apply learning theory to complex human behavior, especially that occuring in real life situations, may have to consider the effects of the variable studied here. This variable might be expected to affect other learning functions as well as the acquisition and extinction of expectancies. It may serve to account for some of the variance usually attributed to human individual differences or chance error in other learning experiments.

Although more research in a wider variety of situations is necessary, the findings of this study and the studies of Phares and Neff also have implications for the clinician who utilizes learning theory as a basis for conducting psychotherapy. In actual practice one may encounter patients who tend to view their experiences both positive and negative as being externally controlled and due to the whims of fate or the manipulation of other people. Before the therapist can either raise or lower expectancies in order to produce behavioral changes, it may be necessary to get the patient to perceive that to some extent the potential reinforcements (positive or negative) in situations are consequences of his own actions and can be controlled. At the other extreme one may encounter patients who attribute most of their experiences to factors they consider to be internally controlled. These patients may blame themselves for frustrating events which were beyond their control and in effect may unrealistically reduce expectancies for gratification.

Summary

This study contrasted 100% and 50% random reinforcement under conditions where *Ss* were instructed that success on a task was controlled by chance or by their own skill. A different group of 20 *Ss* was utilized for each of the four conditions.

The findings revealed, as hypothesized, that under conditions of internal control of reinforcement (skill) the usual superiority of the partially reinforced group in resistance to extinction did not obtain. In fact, under these conditions 100% reinforcement led to less (but not significantly) rapid extinction than 50% reinforcement. The externally controlled groups (chance) were typical of other studies of partial reinforcement. The 100% chance group was significantly faster in extinguishing than the 50% chance group and the 100% skill group. The 50% chance group was significantly slower in extinguishing than the 50% skill group. Some implications of these findings for learning theory and psychotherapy are discussed.

References

1. Dixon, W., & Massey, F. *Introduction to statistical analysis.* New York: McGraw-Hill, 1951.

2. Lasko, A. The development of expectancies under conditions of patterning and differential reinforcement. Unpublished Doctor's dissertation, Ohio State Univer., 1952.
3. Neff, J. W. Individual differences in resistance to extinction as a function of generalized expectancy. Unpublished Doctor's dissertation, Ohio State Univer., 1956.
4. Phares, E. J. Expectancy changes in skill and chance situations. *J. abnorm. soc. Psychol.*, 1957, **54**, 339–342.
5. Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
6. Wyckoff, L. B., & Sidowski, J. G. Probability discrimination in a motor task. *J. exp. Psychol.*, 1955, **50**, 225–231.

(Received June 14, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

UCS INTENSITY AND THE ASSOCIATIVE (HABIT) STRENGTH OF THE EYELID CR[1]

K. W. SPENCE, D. F. HAGGARD,[2] AND L. E. ROSS

State University of Iowa

In a recent series of lectures (7) the senior author tentatively suggested the theoretical possibility that classical aversive conditioning is governed by a reinforcement principle, whereas instrumental appetitional (reward) conditioning does not involve such a principle. Contiguity of the stimulus and behavior events was considered to be a sufficient condition for learning, i.e., habit formation, to occur in this latter type of situation. That is, the increment of habit strength (H) of the instrumental response was assumed to depend only on the occurrence of the response and not to be a function of the reinforcer and its properties. In contrast, the increment of H of the conditioned aversive response was interpreted as being a function of the occurrence and properties of the reinforcing UCS. The present study is concerned with evidence as to whether classical aversive conditioning requires the assumption of a reinforcement principle, or more specifically, whether the habit strength (H) of a defense CR is a function of the intensity of the UCS.

On the basis of the experimental finding that level of performance in classical aversive conditioning varies directly with the intensity of the UCS it has been assumed that drive strength, D, is a function of the intensity (noxiousness) of the UCS. According to a reinforcement interpretation of this kind of learning the habit strength (H) would also be assumed to be a function of the UCS. Thus, according to this conception, performance differences with different intensities of the UCS reflect differences both in H and in level of D.

The experimental problem involved in testing this theory is to find some way of separating out these two factors. A previous study by Spence (6) attempted to accomplish this by employing the factorial design type of experiment. While the results of this experiment were in agreement with the reinforcement interpretation that different amounts of habit strength develop with different intensities of the UCS, as was pointed out, it was also possible to interpret the findings in terms of a differential drive (D) level based on fear responses of different strength that become conditioned to the cues of the experimental situation.

The present studies represent a different type of attack on this problem. In these experiments an attempt is made to equate S's level of drive during the course of the conditioning and at the same time provide for differential reinforcement. In order to understand the logic of these experiments, it is necessary first to consider further the source of drive in classical defense conditioning. In the other main types of conditioning situations, classical and instrumental reward, and instrumental escape conditioning, the drive level (D) is a function of some relevant need produced by the manipulation of a maintenance schedule or by presentation of a noxious stimulus. The need in these instances is active (present) at the moment of the

[1] This study was carried out as part of a project concerned with the role of motivation in learning under Contract N9 onr-93802, Project NR 154-107 between the State University of Iowa and the Office of Naval Research.

[2] Now at Human Resources Research Office, Unit No. 1, Fort Knox, Kentucky.

occurrence of the response and thus provides a certain level of *D* which, in combination with habit strength, determines the excitatory strength of the response. But what determines the need (drive state) and hence the level of *D* in the case of the conditioned anticipatory response on a particular trial in the classical defense type of conditioning? One cannot say that it is the noxious UCS on that particular trial for the CR anticipates this stimulus event. Furthermore, one cannot appeal to a conditioned emotional (fear) response that presumably would be established to the CS. The latency of such a response, mediated as it is by the autonomic nervous system, is much longer than the half-second or less interval of time that has typically been used and found optimal for conditioning such defense reactions. Thus in such instances the conditioned skeletal response, occurring as it does in from 200 to 500 msec., and lasting only briefly, has presumably taken place long before the fear response would even get started.

Because of these considerations it has generally been assumed that the level of drive (*D*) in this type of conditioning experiment is determined by the previous unconditioned stimuli; that is, those given on previous trials. These stimuli, depending upon their intensity, have been assumed to elicit hypothetical emotional responses (r_e) of varying magnitude (**7**). The latter have been thought of as persisting in their effects, lasting at least until the ensuing trial. As a number of writers have pointed out, this persisting emotional response may also get established as a CR to other cues in the general situation, possibly even to *S*'s own verbal responses. These expectations, or fears, are assumed to keep a persisting level of emotional activity present that is, to an important degree, a function of the intensity of the UCS employed in the experiment. We have assumed that the level of *D* is directly related to this level of hypothetical emotional activity.

In the two experiments to be reported the drive level (*D*) of two groups for whom reinforcement was differential was equated by employing two different intensities of the UCS (air puff). For the low-reinforcement group a weak puff always occurred on a conditioning trial, while a strong puff was always presented on a trial on which no conditioning could occur. For the high-reinforcement group these conditions were reversed, the strong puff being presented on the conditioning trials and the weak puff being presented under the conditions in which no learning could take place. Since the two groups received the same average intensity of the noxious puff stimulus throughout the training period, their drive levels would be equal. However, the intensity of the puff on the trials on which conditioning could occur differed for the two groups. According to the present reinforcement interpretation the group that had the strong puff on a trial on which conditioning could occur should develop a greater amount of habit strength than the group that had a weak puff on a conditioning trial. Multiplication of these different *H* values by the same *D* values (since drive level was equal) implies that the excitatory potential (*E*) and, hence, response strength should be greater for the high-reinforcement group than for the low-reinforcement group.

Experiment I

Subjects.—A total of 164 men and women students from an introductory course in psychology were used. Ten of these were eliminated because they met the criteria employed in previous studies in this laboratory defining a voluntary responder (**6, 8**). The data of four other *S*s were discarded, two because they gave CR's in the pretest trials and two for reasons of equipment failure. The remaining 150 *S*s provided three groups, each of which contained 25 men and 25 women.

Apparatus and method of recording.—The *S* was seated in a dental chair in a semidarkened room. The *E* was in an adjoining room in which the stimulus controls and recording equipment were located. The apparatus for recording the eyeblinks and presenting the stimuli was identical with that used in recent studies from this laboratory (see **6**).

The CS consisted of an increase in the bright-

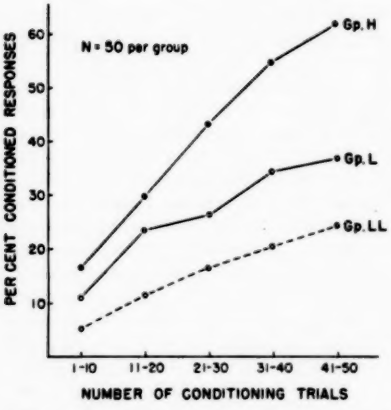

FIG. 1. Acquisition curves showing the percentage of CR's in successive blocks of 10 conditioning trials.

ness of a 6-cm. circular disc from a level of .004 to .506 apparent ft.-candle. The duration of the CS was 500 msec., with the UCS occurring 450 msec. after the onset of the CS. The duration of the UCS, an air puff of either .33 or 2.0 lb./sq. in. applied to the right eye through a .062-in. diameter orifice, was limited to 50 msec., by a 100-V., 60-cycle AC-operated solenoid valve controlled by an electronic timer.

The word "ready" preceded each presentation of the CS by 2, 3, or 4 sec. according to a prearranged schedule. The *Ss* were instructed to blink upon presentation of the ready signal and then fixate the circular disc in front of them. A CR was recorded whenever the record showed a deflection of 1 mm. or more in the interval 200 to 450 msec. following the onset of the CS. Responses with a latency of less than 200 msec., which were infrequent, were classified as original responses and were not included in the data.

Conditioning procedure.—Following the reading of instructions, each *S* received three presentations of the CS alone. A single presentation of the UCS alone was then given. Immediately following these preliminary trials each *S* received 100 trials, 50 conditioning trials which involved the paired presentation after the ready signal of the CS and the UCS and 50 trials which involved the presentation of only the UCS. The order of presentation of these paired and unpaired trials was prearranged according to an irregular order in which the number of each was equalized in blocks of four trials. Intertrial intervals of 15, 20, and 25 sec., averaging 20 sec. and arranged according to a fixed schedule, were used throughout the 100 trials. At the end of the experiment *Ss* were warned not to discuss the experiment with other members of the class.

In the case of the high-reinforcement group (Group H) the CS was always paired with the strong puff, whereas the weak puff was always presented alone, i.e., without the CS. The reverse conditions held for Group L (low reinforcement), the weak puff always being paired with the CS and the strong puff always being presented alone. A third group designated Group LL (low reinforcement and low drive) was also run in which a weak puff was used on both the paired (conditioning) trials and the trials on which the air puff was given alone.

Results.—Figure 1 presents the frequency curves of conditioning for the three groups of *Ss*, 50 in each group. As may be seen, the curve for Group H rises well above that for Group L, whereas the latter is above that for Group LL. Evaluation of the performance levels over the last two points of the curves by means of the Mann-Whitney U test indicated that the difference between Groups H and L was highly significant ($P = .0014$), while the difference between Groups L and LL was significant at the .05 level.

Since the level of D, defined in terms of the intensities of the UCS employed, was equated for Groups H and L, the significant difference in the performance levels of these two groups presumably reflects a differential strength of H. The latter is, of course, related to the different intensities of the puff strengths on the paired conditioning trials. The difference in the curves of Groups L and LL, on the other hand, presumably reflects a difference in the level of D only, since these two groups had equated puff strengths on the conditioning, i.e., H-producing, trials. The difference between Groups H and LL, which was also highly significant ($P = .0002$), would be interpreted by the present theory as reflecting dif-

ferences in both habit strength and drive level.

The findings of this experiment are thus seen to be in accord with those reported in a preliminary experiment in which a regular order of presentation of the paired (P) and unpaired (U) trials (repeated blocks of UPPU) was employed (**7,** p. 177). The irregular order of presentation employed in the present experiment rules out the possibility that the difference between the high- and low-reinforcement groups could have resulted from the conditioning of the *S*'s emotional response to the temporal order.

In order to investigate the possibility that the performance difference between the high- and low-reinforcement groups (H and L) might be due either to a difference in the magnitude of the UCS's made to the different intensities of the UCS on their respective conditioning trials, or to adaptation of the response to the weaker UCS during the conditioning trials, a further analysis of the data was made. Measurement of the amplitudes of the UCR's made to the air puff on the conditioning trials on which no CR occurred over the first 10 and last 10 paired trials revealed the fact that some *Ss* did show considerable adaptation of the eyeblink to the puff, and also that there was a marked difference in Groups H and L in this respect. Accordingly, all *Ss* whose UCR's during the last 10 paired trials averaged less than 50% of their magnitudes in the first 10 paired trials were eliminated. On the basis of this criterion 1 *S* was eliminated from Group H and 11 from Group L, leaving 49 and 39, respectively, in these two groups.[3]

Table 1 gives the mean percentage of CR's made in the last 20 conditioning trials (31–50) by Groups H and L after elimination of these 12 *Ss*. While the difference between the two groups was reduced, it was still significant as tested by the Mann-Whitney U test ($.02 < P < .05$). Thus the performance difference in the case of these *Ss* is shown not to be accountable for in terms of failure of the low-reinforcement *Ss* to respond with adequate UCR's on the conditioning trials.

TABLE 1

MEAN PERCENTAGE OF CR'S IN LAST 20 CONDITIONING TRIALS

Group	N	Mean	σ_M
H	49	59.6	4.43
L	39	43.8	5.06

EXPERIMENT II

Instead of using the UCS alone on the nonconditioning trials the CS was presented on all trials in this experiment. On half of the trials the UCS was administered at a CS-UCS interval (500 msec.) known to produce optimal conditioning, whereas on the other half of the trials this interval was 2650 msec., a duration known to produce little or no conditioning, and to result in extinction when introduced after establishing a CR at a shorter, optimal interval (**2, 3**).

It will be seen that this procedure involved a kind of partial reinforcement for both groups, with half of the trials being reinforced and half nonreinforced. In the case of the

[3] Examination of the amplitudes of the UCR's given by the two curtailed groups during the last 10 paired trials on which no CR's occurred revealed a slight, but insignificant, difference in their mean magnitudes (H = 25.9 mm., L = 24.0 mm.). Within these groups there was no evidence of any relation between the level of conditioning performance and magnitude of the UCR.

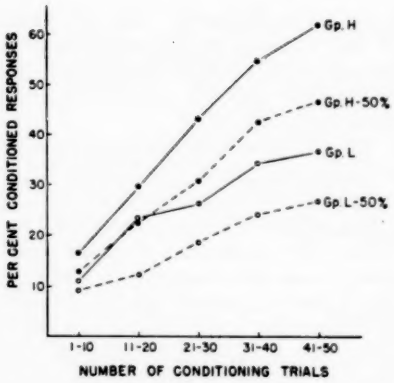

FIG. 2. Acquisition curves showing percentage of CR's in successive blocks of 10 conditioning trials for high- and low-reinforcement groups. The curves for both Exp. I (solid) and Exp. II (broken) are shown in order to reveal the decremental effects of the partial reinforcement schedule used in Exp. II.

high reinforcement *Ss* the reinforcing, *H*-producing trials (short CS-UCS interval) always involved the strong UCS, whereas the weak UCS was present on the nonreinforcing trials. The reverse conditions held for the low-reinforcement *Ss*. According to a reinforcement interpretation the group which received the strong UCS on the short CS-UCS (i.e., reinforcing) trials should show a higher frequency of CR's than the group which received the weak UCS on these trials. Again, since the intensity of the air puffs administered was equated for the two groups during the training period, the performance difference cannot be interpreted as reflecting a difference in *D* and hence would be considered as due to a difference in *H*.

Subjects.—The *Ss* were 47 women and 44 men from an introductory course in psychology. The data of seven *Ss* who were identified by criteria described previously (**8**) as defining voluntary responders were eliminated from the experiment along with that of two *Ss* who gave CR's to initial test trial presentations of the CS alone. Two additional *Ss* were excluded due to *E*'s error. The remaining 80 *Ss* provided two groups, each of which contained 20 men and 20 women.

Apparatus and method of recording.—The experimental apparatus and recording procedures used in this experiment were identical with those employed in the first experiment.

Conditioning procedure.—The instructions and preliminary trials were the same as those used in Exp. I, as was the employment of a 15-, 20-, or 25-sec. intertrial interval, a ready signal which preceded the CS, again an increase in the brightness of a 6-in. circular disk, by 2, 3, or 4 sec. Each *S* received 100 trials, 50 reinforcing trials with a 500-msec. CS-UCS interval, and 50 nonreinforcing trials with a 2650 msec. CS-UCS interval. The order of the two kinds of trials was irregular with the number of each type equalized in blocks of four trials. One group of *Ss* (H-50%) received a 2-lb./sq. in. air puff on the 500-msec. trials and a .33-lb./sq. in. air puff on the 2650-msec. trials, while the other group (L-50%) received the reverse pairing, i.e., a 33-lb. puff on the 500-msec. trials and a 2-lb. puff on the 2650-msec. trials.

Results.—The dotted curves in Fig. 2 present the data obtained in this experiment. They represent the frequency curves of conditioning in terms of the number of conditioning (i.e., reinforcing) trials for the high-reinforcement group (H-50%) and the low-reinforcement group (L-50%). Also presented for comparative purposes are the frequency conditioning curves obtained for high- and low-reinforcement conditions in Exp. I. Corroborating the results of Exp. I in which 100% reinforcement was involved, the curve for Group H-50% exhibited a gradual divergence from that of Group L-50%. Over the last 20 trials the difference in performance level in favor of the high-reinforcement group, tested by means of the Mann-Whitney *U* test, was highly significant ($P < .01$).

The relative position of the conditioning curves obtained in the present experiment with respect to those of Exp. I is of some interest. Unfortunately, through an error in

calibration, the CS-UCS interval employed in the two experiments differed by 50 msec. The interval employed in Exp. I was 450 msec., while that used in the present one was 500 msec. With the exception, however, of this difference and the difference in reinforcing procedure, the experimental conditions in the two studies were the same. Since we know from other unpublished data obtained in our laboratory that a 500-msec. interval produces just as high, if not slightly higher, level of performance than 450 msec., the depressed levels of performance shown by the high- and low-reinforcement groups in the present experiment, as compared with the groups from the first, reflect the inhibitory effects of the nonreinforcing trials.

The findings of this experiment further preclude the possibility that the performance difference in the two previous experiments might have been due to the operation of some kind of differential set with respect to the CS that could have been established in the two reinforcement groups by virtue of the fact that the CS was paired with a UCS of different strength. In the present experiment there presumably was no possibility of such a differential set being present since the CS was followed equally often in both groups by the two different puff intensities. Furthermore, *S* was never aware of which puff intensity would follow the CS on any particular trial.[4]

[4] It is of some interest to note that *Ss* did not acquire a set to respond differentially at the longer interval. Thus, there was no difference between the two groups of the present experiment in the number of anticipatory responses occurring 1 sec. prior to the UCS in the long-interval trials. Such responses are, for the most part, random blinks, for we know that there is little, if any, conditioning at this long interval.

As in Exp. I, an analysis of the data was also carried out in which *Ss* who exhibited adaptation to the UCS were eliminated. Thus all *Ss* whose UCR's during the last 10 reinforcing trials averaged less than 50% of the magnitudes in the first 10 reinforcing trials were eliminated. In the case of the high-reinforcement group these were UCR's made to the 2.0-lb. puff, whereas in the case of the weak-reinforcement group they were responses given to the weak (.33 lb.) puff. On the basis of the criterion of elimination, one *S* was eliminated from Group H-50% and nine from Group L-50%, leaving 39 and 31 *Ss*, respectively, in the two groups.

Table 2 shows the mean percentage of CR's made in the last 20 conditioning trials (last 40 trials) by the two groups of *Ss* after elimination of the 10 adapters. Again, as in the previous experiment, the difference between the two groups was reduced, but was still significant at the .05 level. That there was no relation within these *Ss* between the amplitude of the UCR made to the reinforcing UCS and the percentage of CR's is shown by the fact that the correlation coefficient between them was −.046. In the case of all *Ss* this coefficient was .245.

This finding thus confirms the previous experiment in showing that the performance difference between the two groups is not due to a failure of the low-reinforcement *Ss* to respond

TABLE 2

MEAN PERCENTAGE OF CR'S IN LAST 20 CONDITIONING TRIALS

Group	*N*	Mean	σ_M
H-50%	39	47.8	5.4
L-50%	31	31.6	3.9

with an adequate UCR on the conditioning trials. Perhaps attention should also be called here to the point that the failure of the *Ss* whose UCS adapts out to show conditioning is quite in accord with a reinforcement interpretation that learning (i.e., growth of *H*) is a function of the intensity of the reinforcing aversive stimulus. Such *Ss*, according to this interpretation, should show relatively little, if any, conditioning.

Discussion

The evidence of these two experiments and the similar experiment previously reported by Spence (**7**, p. 177) shows clearly that, with level of drive (*D*) equated, performance in classical aversive conditioning is a function of the intensity of the UCS occurring on the reinforced trials. This finding may be interpreted as lending support to a reinforcement-type theory that habit strength (*H*) in such aversive conditioning is a function of the intensity of the UCS. One version of such a reinforcement theory is the drive-stimulus reduction conception of Hull (**1**) and Miller and Dollard (**4**). According to this view the cessation of a strong UCS in aversive conditioning would provide greater reinforcement than cessation of a weak UCS and thus should lead to a greater increment of *H* per trial. The present writers would prefer to confine their interpretation to the more general conception that habit strength (*H*) is some function of the intensity of the UCS, leaving the nature of the reinforcing mechanism out of consideration.

In connection with the decremental effects of the partial reinforcement procedure employed in Exp. II, attention should be directed to the evidence provided in Fig. 2 by a comparison of the upper two curves that the inhibitory effect developed gradually and consistently over the 50 trials of conditioning and 50 nonreinforced trials. In contrast, the two lower curves suggest that the inhibitory effect reached a maximum much earlier in the low-reinforcement condition, possibly as soon as 20 conditioning trials or a total of 40 trials. Further data providing such comparisons are needed so that we may ascertain the nature of the development of such inhibitory effects on performance level during conditioning under partial schedules of reinforcement.

Summary

This study was concerned with the problem of whether habit strength (*H*) is a function of the intensity of the UCS in classical aversive (eyelid) conditioning. In Exp. I one group (high reinforcement) always had the CS paired with a strong UCS and in the other group (low reinforcement) with a weak UCS. The drive (*D*) level was equated (on the average) in the two groups by interspersing among the conditioning trials an equal number of trials with the UCS alone. Thus the high-reinforcement group was presented with the weak UCS on such trials and the low-reinforcement group the strong UCS. Comparison of the frequency of CR's in the last 20 conditioning trials revealed a significant difference between the high- and low-reinforcement groups. This result was shown to hold even with *Ss* equated for the magnitudes of their UCR's.

In Exp. II one group (high reinforcement) had the CS, a light, paired on half the trials with a strong UCS at an optimal CS-UCS interval (500 msec.). On the other half of the trials the CS was paired with a weak UCS at an interval (2650 msec.) known not to lead to conditioning. The conditions for low-reinforcement group were the reverse, the CS being paired with the weak UCS on the conditioning (i.e., reinforcing) trials and the strong UCS on the nonconditioning trials. Performance measured over the last 20 conditioning trials (40 total trials) revealed that the high-reinforcement group gave a significantly greater number of CR's than the low-reinforcement group.

Since the drive levels of the high- and low-reinforcement groups in each experiment were equated, the performance differences between them were interpreted as reflecting a difference in the learning factor (*H*) and hence as supporting a reinforcement type of learning theory as far as aversive conditioning is concerned.

References

1. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

2. McAllister, W. R. The effect on eyelid conditioning of shifting the CS-US interval. *J. exp. Psychol.*, 1953, **45,** 423–428.
3. McAllister, W. R. Eyelid conditioning as a function of the CS-US interval. *J. exp. Psychol.*, 1953, **45,** 417–422.
4. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.
5. Passey, G. E. The influence of the intensity of unconditioned stimulus upon acquisition of a conditioned response. *J. exp. Psychol.*, 1948, **38,** 420–428.
6. Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45,** 57–63.
7. Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.
8. Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.

(Received March 4, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

A FURTHER TEST OF THE HYPOTHESIS OF AUTONOMOUS MEMORY TRACE CHANGE

S. H. LOVIBOND

University of Adelaide

The Gestalt hypothesis of autonomous change in the memory trace has important implications for theories of learning and perception based on neuro-physiological mechanisms. Thus Hebb (**5**) found it necessary to reject the evidence favoring this hypothesis in order to clear the ground for his postulation of a structural trace as the basis of memory. However, various methodological objections have been raised against *all* investigations so far designed to test the Gestalt hypothesis, irrespective of the findings.

The criticisms of the earlier work based on the method of repeated reproduction have been well summarized by Hebb and Foord (**6**) who state a number of criteria for an adequate test of the hypothesis. In Hebb and Foord's own study, and a later experiment by George (**3**), the method of recognition was used, with different groups of *Ss* at each retention interval. The results of both experiments were negative.

In a later study, Carlson and Duncan (**2**) followed the general procedure of Hebb and Foord, but introduced different stimulus figures. The results obtained by Carlson and Duncan were also inconsistent with the Gestalt hypothesis, but these authors regarded their own experiment, and those of Hebb and Foord, and of George, as inconclusive on the grounds that the recognition tests employed may have distorted the results. They noted a tendency for *Ss'* judgments to err in the direction of the error contained in the recognition figure presented initially.

Although Carlson and Duncan do not draw the conclusion, it would appear that the method of recognition which requires *S* to make a selection from a temporal series of recognition figures, makes use, in effect, of a series of repeated recognition tests. As a result, it suffers from defects similar to those of the repeated reproduction method, i.e., *S's* judgment is based on the interaction of the trace of the original test stimulus and the traces of preceding recognition figures. This suggests that, in order to avoid such an effect, the method of recognition should employ only a single recognition test per *S*. The method then becomes the "method of comparison" used by Hanawalt (**4**). Hanawalt has criticized this method on the grounds of instability of results and the masking of true forgetting. It is employed in the present investigation with a number of modifications: (*a*) it is used in conjunction with a single reproduction test; (*b*) the test stimulus is not the original standard stimulus but a replica of a previous "best estimate" of the standard; and (*c*) *Ss* are unaware that a later memory test is to be made.

Apparatus

Stimulus figure.—The stimulus figure used was the inverted **V** of Carlson and Duncan. Two inverted **V**-shaped apertures, 38 cm. apart at the apexes, were cut in the front sheet steel wall of a box approximately 60 cm. by 28 cm. The angle of the **V**'s was 65° and the width of each arm was 1.4 cm. The left leg of each **V** was 18 cms. in length. The right leg of the

left-hand V was made continuously variable in length from zero to 20 cm. by means of a slide actuated through a chain and pulley mechanism. Adjustment was effected by rotation of a wheel on the side of the box. The setting could be read to the nearest millimeter on a ruler mounted parallel to the adjustable leg and traversed by a pointer attached to the slide. Each V aperture was illuminated independently through semi-opaque plastic fastened to the inside surface of the front wall of the box.

Subjects.—The *Ss* were 120 male and female adults. Most were first-year psychology students. They were divided into 6 groups of 20 each, 1A, 2A, 3A, 1B, 2B, and 3B. For all A groups the short leg of the standard V was 13.5 cm. in length. For the B groups the standard leg length was 12 cm. The retention interval was 3 min. for Groups 1A and 1B, one week for Groups 2A and 2B, and two weeks for Groups 3A and 3B. These retention intervals paralleled those of Carlson and Duncan.

Procedure.—The *Ss* were brought individually into a darkroom and seated facing the apparatus at a distance of 8 ft. The *S* was told that his task was to instruct *E* how to adjust the right-hand V in order to make it exactly the same as the V on the left. Four trials were given using the method of adjustment. Two trials began with the adjustable leg longer than the standard and two began with the adjustable leg shorter than the standard. The mean of the four adjustments made by *S* was computed and *S* was asked to leave the room and wait or return one or two weeks later.

As in Carlson and Duncan's experiment, *S* was not told that memory was involved in the experiment. At no time was the stimulus figure referred to other than as a "patch of light."

In the test of retention, *the comparison V was adjusted to give a leg length equivalent to the mean of the S's judgments in the first part of the experiment.* The *S* was brought into the room and reminded of his previous task. He was told that this time he would see only the comparison V which had been set to a predetermined setting. This setting differed for the different groups of the experiment and his task would be to instruct *E* which way to adjust the comparison V in order to make it exactly the same as the V on the left as he remembered it.

The comparison V was then switched on and *S* was asked to make his judgment. If he stated that no change was necessary, a choice was forced. After the initial judgment of direction was made, *E* adjusted the V in the chosen direction until *S* was satisfied that it now matched the standard as far as he could remember it. The setting was then read off the millimeter ruler and recorded.

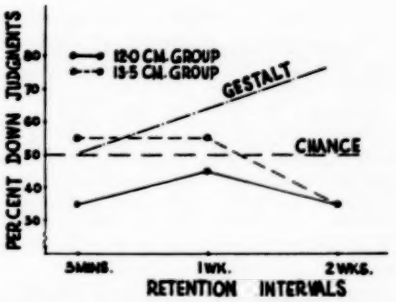

FIG. 1. Results of recognition test in percentage of down judgments at each retention interval.

Finally, open-ended questioning followed by funneling was used to ascertain the extent to which *S* used verbal aids to recall or relied on a visual image. The responses were categorized as "visual only" ("I simply saw it in my mind's eye and compared it with this one") and "other" ("I remembered the right-hand side of the figure was about ¾ of the length of the other one and I saw at once that the one you showed me had the short side about ¾ as long so I knew it had to go up").

RESULTS

In the first part of the experiment there was a significant tendency for *Ss* to set the adjustable leg of the comparison V too short. The mean settings of 96 *Ss* were shorter than the standard while the mean settings of only 14 *Ss* were longer than the standard. In the remaining 10 cases the setting was identical with the standard.

Figure 1 shows the results of the recognition test in terms of the percentage of down judgments for each group at each retention interval. Also shown are the trends predicted by the Gestalt hypothesis and the hypothesis of chance determination of judgments.

There was no tendency for judgments to be made in the direction

TABLE 1

ERRORS IN JUDGMENT (IN MILLIMETERS)

Retention Interval	Group A		Group B	
	Mean	*SD*	Mean	*SD*
3 min.	+0.45	2.1	−1.5	1.09
1 week	−0.10	6.9	+5.25	1.52
2 weeks	−2.65	6.3	−2.25	1.11

of the standard. Of the 110 *Ss* whose original mean settings contained error, 56 requested an adjustment in the direction of the standard, and 54 gave a judgment away from the standard.

The data were tested for trends by the extension of the Mann-Whitney test proposed by Bross (**1**) for dichotomies with ordered categories. For Group A, $t^2 = 1.57$ and $P = .22$, and for Group B, $t^2 = 0$ and $P = 1.0$. These values indicate that the hypothesis of chance determination of judgments cannot be rejected.

The results of the reproduction test are presented in Table 1 which contains the means and *SD*'s of the errors made by each group at each retention interval.

In this case a test of the hypothesis of zero trend against the alternative of an ordered sequence of change with time was made by means of the test proposed by Jonckheere (**7**). This test is a further extension of the nonparametric Mann-Whitney test.

For Group A, $S_o = 1.48$ and $P = .14$, while for Group B, $S_o = .1$ and $P = .92$. These figures also are consistent only with the hypothesis of chance determination of judgments.

Table 2 gives the frequency with which visual and other means of recall were employed by the 98 *Ss* from whom these data were obtained.

The table shows that the majority of *Ss* in each group relied on visual imagery only. This would seem to indicate that the procedure used was successful in minimizing verbal factors in recall.

DISCUSSION

It is clear that the data offer no support for the Gestalt hypothesis of slow unidirectional change in the memory trace. Hence it becomes appropriate to raise the question whether conditions favorable for the manifestation of such a phenomenon have been provided.

It would appear that this question can be answered in the affirmative. In addition to meeting the criteria for an adequate test of the hypothesis laid down by Hebb and Foord (**6**), the present procedure has minimized verbal influences and precluded conscious effort to memorize by presenting the task as one of visual discrimination.

In those experiments which include specific instructions to remember the stimuli for the purpose of a later memory test, *S* usually makes every effort to "lay down a stable trace," typically by resort to verbal aids. For example, George (**3**) found that his *Ss* tended to remember verbal phrases such as incomplete circle, black ellipse, etc. Although according to Gestalt theory the process of autonomous change ordinarily tends toward dominance over all influences external to the trace, it could be argued, nevertheless, that under these special conditions verbal factors might mask the appearance of autonomous change.

TABLE 2

FREQUENCIES OF DIFFERENT METHODS OF RECALL IN THE TWO MAIN GROUPS*

Method	Group A	Group B	Total
Visual Only	36	37	73
Other	12	13	25
Total	48	50	98

* These data were not obtained from all *Ss*.

Finally, it seems reasonable to conclude that the procedure afforded a sensitive, unbiased measure of the occurrence of autonomous change. The method of using *S*'s "best estimate" of the standard as the base line in the test period effectively removed pre-memory matching errors. Hanawalt (**4**) has objected that the method of comparison is unstable and masks true forgetting.

It may well be that true forgetting in the form of *passive decay* of the trace is masked by the introduction of the stimulus figure (or, as in the present experiment, a best estimate of the stimulus), but this type of trace change is not in question. On the other hand, it is difficult to see how systematic unidirectional change could effectively be masked by this procedure. Certainly it seems likely that the perceptual effect of a stimulus will interact with the trace formed by a previous presentation of that stimulus and tend to remove discrepancies in the trace. Similarly it is probable that a stimulus differing but little from the original stimulus (say in the single dimension of length) will tend to change the trace in the direction of congruence with that stimulus.

If this assumption is correct, *any* method which introduces in the test period either the original stimulus or one akin to it may be expected to produce this masking effect. This would apply not only to the method of repeated recognition trials, but also to the reproduction method when the reproduction is "built up" as in the drawing of a line.

However, in order to account for the present complete lack of evidence for unidirectional change in these terms, it would be necessary to assume that the masking effect is strong enough to reduce to the same level traces which are discrepant by differing amounts. Such as assumption would imply that the trace is rather more fragile than appears to be the case from the available evidence.

Summary

The Gestalt hypothesis of autonomous change in the memory trace was tested by a combination of the methods of recognition and reproduction. The stimulus figure used was an illuminated inverted V with a shortened right leg. Separate groups of *S*'s were used at retention intervals of 3 min., 1 week, and 2 weeks. The results showed no evidence of a progressive unidirectional change in the trace although it was concluded that conditions favorable for the manifestation of this phenomenon had been provided.

References

1. Bross, I. D. J. Is there an increased risk? *Fed. Proc.*, 1954, **13**, 815–819.
2. Carlson, J. B., & Duncan, C. P. A study of autonomous change in the memory trace by the method of recognition. *Amer. J. Psychol.*, 1955, **68**, 280–284.
3. George, F. H. Errors of visual recognition. *J. exp. Psychol.*, 1952, **43**, 202–206.
4. Hanawalt, N. G. The method of comparison applied to the problem of memory change. *J. exp. Psychol.*, 1952, **43**, 37–42.
5. Hebb, D. O. *The organization of behavior.* New York: Wiley, 1949.
6. Hebb, D. O., & Foord, E. N. Errors of visual recognition and the nature of the trace. *J. exp. Psychol.*, 1945, **35**, 335–348.
7. Jonckheere, H. R. A distribution free K-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

EFFECTS OF WINDOW SIZE AND TWO CONDITIONS OF JUDGMENT ON THE REPRODUCTION OF OBJECTIVE VELOCITY[1]

OLIN W. SMITH

Cornell University

In Brown's 1931 review of perceived velocity, he reported the velocity transposition phenomenon: "If a moving field in a homogeneous surrounding field is transposed in its linear dimensions as 1:2 the stimulus velocity must be transposed by a like amount in order that the phenomenal velocity in both cases be identical" (**1**, p. 211). However, Brown's discovery was not exactly a transposition phenomenon, as his own experiments demonstrated. For instance, if Field A in all respects was twice the linear dimensions of Field B and was observed at the same distance as B, the physical velocity of A had to be twice that of B for the apparent velocities of the two fields to be judged equal. But he also obtained matches of equivalent physical value when Field A was observed at *twice the distance* of B. This indicated that his effect, in part, might be due to a difference in window size in which the translatory motion was observed.[2]

Brown's apparatus was approximated at Cornell, and a series of exploratory experiments were made. In one case the size of the field and distance of observation of Field A were twice that of B with the same moving patterns. Each field was surrounded by a fine-grained, plain brown paper. When the physical velocities of both fields were equal, the movement in the small field, B, appeared to be unequivocally faster. When the fields were the same size and speed, and one field either the same or twice the distance of the other, no apparent velocity difference was observed. This proved that a difference in distance and its corresponding difference in angular velocity did not cause the effect.

Differences in degree of apparent velocity were caused by variations in the size of the windows in which translatory motion was observed. However, it is possible that a different set of instructions could cause *O* to make quite different judgments of speed of movement. Brown claimed to be investigating "phenomenal velocity" which is quite different from judgments of objective (physical) speed as required by many tasks such as driving a car or flying a plane. The purpose of Exp. I was to determine whether size of windows had measurable effects on judgments of objective velocity.

The results of laboratory experiments and the data of daily behaviors may differ considerably. One cause of such differences is that the methods of obtaining data in the laboratory may be optimal for minimizing errors of performance while similar behaviors in daily life have higher error components attributable to less favorable conditions for judgment and performance. For that reason, two conditions for velocity matches were compared in Exp. II. The optimal

[1] This paper was supported in part by Contract NONR 401(14) between Cornell University and the Office of Naval Research, Department of the Navy, Washington, D. C. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Translatory motion is a physical, not a psychological, concept and denotes a field or body in motion in the same direction at the same velocity. In terms of visual angles per second, the retinal stimulation resulting from such motions can be, at best, only an approximation of translatory motion.

TABLE 1

VISUAL ANGLES SUBTENDED BY THE WINDOWS

Distance of Observation	Horizontal		Vertical	
	Large	Small	Large	Small
6 ft.	11°17′31″	5°40′00″	33°37′12″	17°10′44″
12 ft.	5°40′00″	2°49′22″	17°10′44″	8°38′20″

conditions of Exp. 1 in which *O* could alternate his view of the standard and variable at will were compared with less favorable conditions in which *O* made a series of matches (from memory) after the standard had been stopped.

A velocity match by *O* may be considered not only as evidence of the effects of the given conditions, but also as a test of *O*'s ability for performing the task. For practical purposes individual differences in ability may be at least as important as differences in conditions for performance. Consequently, the consistency of individual differences is investigated in Exp. II by correlating performance scores under optimal and less favorable conditions for reproduction.

EXPERIMENT I

Method and apparatus.—Two identical motion devices were used (**4**). In each, an endless rubber belt revolved around a pair of cylindrical aluminum rollers which were driven by a variable speed drive and mounted in angle iron frames. A flat vertical surface nearly 4 ft. high and 3 ft. wide could be exposed to observation through a plate glass window mounted on the frames, the dimensions of which could be varied by taping a paper shield to the glass. By pushing a toggle switch, *O* could accelerate or decelerate the variable surface to obtain a match with the standard. The *E* operated a small switching station which could cut *O*'s control box in or out, or control the velocity of either moving surface. The moving pattern consisted of a checkerboard of 1 cm. squares.

Paper windows of equal outside dimensions completely covered the fronts of both machines and could be switched readily so as to change the size of the window of either movement field. The outside dimensions of the "surround" of both movement fields were constant. Each apparatus was illuminated by four 40-w., cool light, fluorescent tubes mounted on either side and parallel to the long axis of the windows. A metron tachometer was used to measure physical velocities in feet per minute. The *O* had a clear, uninterrupted, binocular view of the floor extending from his fixed chair to each apparatus.

Design.—The method of reproduction was used. The *O* adjusted the speed of the variable belt to match the objective speed of the standard belt. The standard was always on the left at 12 ft. from *O* while the variable was always on the right at 6 ft. The linear speed of the standard surface was constant at 24 ft. per min., and *O* had to turn his head 90° to compare surfaces. Space errors were uncontrolled since Brown had found none, and no significant space errors were found in a preliminary study at Cornell.

In Cond. I the large window (110.50 × 36.20 cm.) was used for both the standard and the variable fields. In Cond. II, both fields were observed with the small window (55.25 × 18.10 cm.). In Cond. III, the small window was at the standard, and the large window at the variable, The window sizes in Cond. IV were the reverse of Cond. III. The *O* made four judgments of objective velocity matches for each of the four conditions. Table 1 gives the visual angles subtended by the windows of two sizes at the two distances. The order for each series of 16 judgments was drawn from a table of random numbers for each *O*.

Conditions I and II provide for two estimates of what has been called motion constancy. The mean matches of the four conditions provide estimates of the effects of all possible combinations of the two window sizes.

Procedure.—Each *O* was given instructions and practice in controlling the speed of the variable motion field. For each *O* the prejudgment settings of the speed of the variable were in an ABBA order at 0 ft./min. or 50 ft./min. The standard was then set at 24 ft./min. The *E* instructed *O* to look first at the standard and then at the variable, and to adjust the speed of the variable "so that its velocity in feet per second would be the same as that of the surface on the left." He was told not to use any supplementary measuring scheme, and not to make his match by counting, timing, breathing, or by any other trick or computational aid. The *O* was cautioned not to make his matches by alternately fixating the edges, corners, or points within the move-

TABLE 2

MEAN OBJECTIVE VELOCITY MATCHES FOR THE FOUR CONDITIONS

Cond.	Velocity (Ft. per Min.)				
	Mean	*SD*	σ_m	CE	CE/σ_m*
I	30.09	6.89	1.47	+6.09	4.1457
II	28.31	4.29	0.91	+4.31	4.7119
III	32.52	7.45	1.59	+8.53	5.3702
IV	26.91	4.40	0.94	+2.91	2.6756

* All significant at $P = .001$.

ment fields. (In an exploratory experiment, exceedingly accurate matches were made when *O* judged the rate that objects passed a fixated point.) The *O*'s match was to represent his best judgment of velocities which would be equal *if measured in terms of feet per second*. He could inspect alternately either field as long as he wished. When satisfied with his match, *O* was to inform *E* and then close his eyes until instructed to open them. The *O* was never corrected or informed of his errors.

Subjects.—The *O*s were 22 paid graduate or undergraduate students with at least 20/30 corrected vision.

Results.—The predictions of differences between means of the conditions were based on the rationale that, other things being equal, the apparent velocity in a small window would be faster than the apparent velocity in a larger window. These predictions were:

(*a*) The mean matches of Cond. I (large windows) and Cond. II (small windows) should be the same. Brown's data show no basis for predicting a difference.

(*b*) The mean match of Cond. I (large windows) should be less than that of Cond. III in which the small window was at the standard, far, and the large window was at the variable, near (see Table 2). The apparent velocity of the standard in Cond. III should be faster than that of the standard in Cond. I. In both conditions the variable was observed through the same large window, so no difference between mean matches of the two conditions can be due to window size of the variables.

(*c*) The mean match of Cond. I should be greater than that of Cond. IV. The windows of the two standards were the same (large). Two sizes were used for the variable with the smaller window in Cond. IV. The apparent velocity in the small window should be greater than that in the large, so that the physical value of its match should be less. The same rationale decided the remaining predictions as marked by two asterisks in Table 3.

Tables 2 and 3 list the results which are in the predicted direction in five out of the six predictions. The obtained difference between means of Cond. I and II was contrary to the prediction of no difference, but it was not significant. The five remaining predictions were in the correct direction and three of these were significant at the 1% level. The hypothesis thus remains tenable that window size affects judgments of objective velocity. When the window size of the variables was the same, but those of the standards were unequal, differences were not significant (see Table 3, Cond. I vs. III, II vs. IV). When the window sizes of the stand-

TABLE 3

t VALUES OF DIFFERENCES BETWEEN MEAN MATCHES

Cond. Compared	Differences	*t*	*P*
I –II*	+1.78 ft./min.	1.655	NS
I –III**	−2.44 ft./min.	1.757	NS
I**–IV	+3.18 ft./min.	3.408	<.01
II –III**	−4.22 ft./min.	9.145	<.01
II**–IV	+1.40 ft./min.	0.521	NS
III**–IV	+5.62 ft./min.	4.004	<.01

* No difference was predicted for this pair.
** The mean predicted to be the largest of the two.

TABLE 4

MEANS AND VARIANCES OF MATCHES IN SERIES A AND B

Match No.	Series A		Series B	
	Mean	Variance	Mean	Variance
1	30.20	35.75	30.50	52.16
2	28.40	21.41	30.45	88.26
3	28.75	31.14	31.40	102.36
4	30.40	34.15	31.15	41.61

ards were the same, but those of the variables were unequal, differences between pairs of means were significant (see Table 3, Cond. I vs. IV, II vs. III). Window size of the variable field was more important than that of the standard, but this conclusion is restricted to comparisons in which window size differences exist within one of the pairs being compared.

The *SD*'s of Cond. II and IV are smaller than those of Cond. I and III. In each case the smaller *SD* was obtained when a match was made in the small window of the variable. The *t* tests of the difference in variance between Cond. I–II, I–IV, and II–III were significant ($P < .01$) while the *t* for Cond. III–IV was significant only at $P < .02$. Accuracy was greater in the small windows than it was in the large windows of the variable.

The mean matches of all four conditions were significantly greater than the velocity of the standard.

EXPERIMENT II

Method.—The conditions, procedures, apparatus, and instructions were those of Exp. I except as modified below. Twenty of the 22 *O*s of Exp. I were assigned at random, five each, to the four combinations of window sizes. Four matches were made (Series A) during which *O* alternated his view of standard and variable at will. After the fourth match the standard was stopped, its lights extinguished, and *O* was instructed to make four additional matches from memory. He was to adjust the velocity of the variable so that it would be equal in terms of ft./sec. to that of the standard which he had been observing. The *O* was instructed before the fourth judgment that it would be his last. After each match *O* looked aside or closed his eyes while the velocity of the variable was measured.

Results.—The mean matches of Series A and B were 29.44 and 30.88 ft./min. (the score per *O* being the mean of his matches per series). The *t* test of their difference was not significant. The variances of Series A and B were 19.0123 and 50.2401, and the *t* of the difference was significant ($P < .01$). *t* tests were then made of the differences between serially adjacent means and variances (score per *O* being his raw score per match. See Table 4). No difference between serially adjacent means or variances was found that was attributable to an order effect. The variance of Match 4, Series B, was significantly smaller than that of Match 3 ($P < .02$). This is attributed to the instruction that Match 4 would be the last of the series. Reliability estimates were computed between raw scores per *O* of serially adjacent matches (see Table 5). The reliability estimates of Series B were uniformly greater ($P < .05$) than these of Series A when tested by paired replicates (6). The corre-

TABLE 5

RELIABILITY ESTIMATES

No. of Match	Series A No. of Match			Series B No. of Match		
	1	2	3	1	2	3
2	.52**			.52**		
3	.42	.40		.58*	.69*	
4	.58**	.41	.23	.71*	.64*	.72*

* Significantly different from zero at 1% level.
** Significantly different from zero at 5% level.

lation between the last match of Series A and the first match of Series B was .73. The correlation between the mean match per *O* for the two series was .67.

Discussion

Brown's goal was to ". . . find the factors that condition phenomenal velocity" (**1**, p. 207). In so doing he asked his *O*s to judge which of two moving objects, as viewed successively, was moving faster. The above instructions require the assumption that stimulus movement induces a dominant percept, immediate in experience, and discriminable from a second such experience by virtue of an attributable velocity difference.

Another assumption is equally tenable. The instruction caused *O* to judge whatever dimension(s) of the total stimulus array may be judged *most easily* within a commonly accepted connotation of the word "faster." The immediate experience, however, may be the same in either case. This latter assumption necessitates some discriminable difference in stimulus dimension(s) to be in correspondence with an accepted meaning of "faster." Either judgments of faster based on discriminable differences in the stimulus dimension(s) for objective velocity, or judgments of other stimulus dimension(s), would have been completely in accord with Brown's instructions. Stimulus dimensions are described below which are in correspondence (*a*) with judgments of objective velocity, and (*b*) with Brown's data.

Stimulus dimensions for "phenomenal velocity."—In one of Brown's experiments the linear dimensions in all respects and distances of observation of Field A were twice those of Field B. Brown instructed his *O*s to tell which of the two moving objects, when viewed successively, was moving faster. Judgments of *any one of several stimulus* dimensions would have yielded his data. For instance, judgments of the completion of the traverse of the extent of the small field by a point or set of points at the same time that a corresponding point or set of points completed a traverse of the large field would have done so. The simple and easy judgment of first or quicker would have sufficed, and would be equivalent to Wallach's proposal: "Phenomenal speeds are equal when the displacements per unit of time are the same in proportion to the dimensions of their respective fields" (**4**, p. 548).

A judgment of "faster" would be in correspondence with the number of j.n.d.'s in change of position (movement) per unit of time. Duncker (**3**) investigated perceived movement and proved that judgments of movement were dependent upon a change in position of one object relative to another object. Cartwright (**2**) demonstrated that localization of the center of a small rectangle is much more accurate than the localization of the center of a larger rectangle. The results of the present study prove that accuracy of reproductions of movement is significantly better in small windows of the variable than in large ones. The j.n.d.'s for static localization as well as for movement reproduction appear to be a function of the size of the field or frame of reference. If the number of j.n.d.'s of change in position per unit of time is an approximate Weber function, and dependent upon the size of the field, judgments of such changes should be in close agreement with Brown's data. For instance, the physical value of the match in the small field, A, would be about half that of Field B, and these were the values obtained by Brown. However, it should be noted that when judgments of different stimulus dimensions of a constant stimulus array, the standard, can be made in accord with one set of instructions with the logical expectancy of equal results, the phenomenal method as used by Brown is useless for detecting the specific stimulus dimension(s) which are being judged.

Stimulus dimensions for judgments of objective velocity matches.—Refer again to Fields A and B as described above. A perfect judgment of equality of objective velocity should result should *O* judge correctly the time of traverse of the small window by a point and the time for a corresponding point to pass through an equal extent in the large window. A necessary condition for this judgment is that the length of the small window be correctly estimated along the length of the large window. The surround of the movement fields in this experiment probably made the latter judgment more difficult than if a point along the edge of the large field could have been easily identified after having been judged as marking an extent equal to that of the small window.

Reasons for the judgments of Brown's Os.—Although it is impossible to specify the exact stimulus dimensions which were judged by Brown's *O*s, the judgment of which of two points

completed the traverse of its respective field first was probably simpler and easier than the judgment of the stimuli for objective velocity. It seems unlikely that *O*s would choose the more difficult of two judgments which satisfy an instruction equally well. Secondly, if the j.n.d.'s of change in position per unit of time (movement) are a function of field size, a difference based on rate of j.n.d.'s should be more easily judged than the stimuli for objective velocity.

Motion constancy.—The statement of the motion-constancy problem follows the model of the traditional size-constancy problem. Perfect motion constancy is obtained when the physical match of a variable and a standard are equal for two identical movement fields observed at different distances. The problem is to explain why equal physical velocity matches of two movement fields can be made when the angular retinal velocities of the two fields are different. Wallach discussed the problem extensively and proposed an ingenious solution dependent upon only the retinal velocities of the images (**5**). However, the generality of such a solution in terms of retinal velocities alone, seems questionable.

Brown demonstrated that phenomenal matches of equality are made not only when the angular velocities of two fields are the same, but also when one, A, is twice that of the other, B (**1**). In one experiment both fields were observed at the same distance, while in another, A was at twice the distance of B. The mean physical matches were equal under both conditions of observation for judgments of phenomenal velocity. When two different retinal velocities are the proximal stimuli for two matches, equal in retinal velocity, a solution in terms of retinal velocities alone cannot be adequate.

Judgments along some stimulus dimension(s) will yield zero-constancy data, while other dimensions will yield perfect constancy. For example, for an *O* instructed to judge proportional extents of traverse of corresponding points through two fields of different size, zero constancy results are predicted for ideal viewing conditions. If, on the other hand, *O* is required to judge objective velocities as equal under such circumstances, perfect constancy matches are predicted subject to certain limitations.

The *E* trained himself to judge objective velocity under the four conditions of this study. Window effects were eliminated by the training, and constant errors shrank to values less than the unit of measurement, 1 ft./min. For this one trained *O*, objective matching—that is, motion constancy—was nearly perfect. Perhaps the consistency of Brown's data can therefore be attributed to the ease of discrimination of the aspect or aspects of the stimulus array judged by his *O*s.

The objective velocity matches of the present study were made by untrained *O*s. In view of the successful training of one *O*, the demonstrated window effects as well as the differences between the mean matches and the value of the standard may be attributed to the partial learning status of the *O*s which facilitated errors of judgment and discrimination. In this sense, psychophysics and perceptual learning may be intimately related. If the investigator does not have a satisfactory index of the learning status of his *O*s in the performance of the task under study, it is likely that his psychophysical data will be only pertinent to the points on the learning curve which he has sampled. This is more likely to be the case in strange or complex stimulus conditions than for some tasks such as weight lifting where the learning to discriminate and report in accordance with instructions can proceed at a very rapid rate.

Comparison of matches made directly and from "memory."—The variances of the optimal and less favorable conditions of Series A and B were significantly different although their means were not. This indicates that the central tendency of the judgments under both conditions was equivalent, but that the less favorable conditions produced more variability of matching. The greater variance of Series B is not a necessary consequence of the method, since the variance of Match 4 is not significantly different from the matches of Series A. It is assumed that the decrease in variance on Match 4 was due to increased effort, motivation, or set for attending, following the instruction that this was to be the last match of the series.

Exactly what was being "remembered" in Series B was not as evident as in Series A since there are several hypotheses of nearly equal face validity. Some of these are: (*a*) the dominant (functionally most effective) "memory" was that of the standard for the last match of Series A. (*b*) The dominant memory was that of a mean of the memories of the standard in Series A. (*c*) The dominant memory of each of Matches 2, 3, and 4 was that of the memory of the preceding match. (*d*) The dominant memory was an interaction of the above three.

Hypothesis *a* is tenable, since no mean of Series B is significantly different from the mean of Match 4, Series A. Hypothesis *b* is also tenable since the means of the mean match per *O* were not significantly different for the two series. Hypothesis *c* is untenable according to the following rationale. If each match in Series B is based on the preceding match in the series, the constant error of overestimation obtained under optimal conditions should be

added in turn to increase each successive match. Therefore, there should be *at least* a significant difference (which was not obtained) between the means of Matches 1 and 4. The difference between means of Series B are so small that it is reasonable to believe them anchored by a constant standard throughout the series, rather than by shifting standards. Hypothesis *d* remains tenable since the data are not adequate for its test.

Reliability.—The low reliability estimates of Series A are consistent with the assumption that a judgment of equality under optimal conditions is an estimate of the center of the range of uncertainty in which *O*'s judgments can be expected to be randomly distributed around his mean. The higher correlations of Series B are attributable to the greater range of matches, 14–61 ft./min. as compared with 18–41 ft./min. for Series A. The best estimate of reliability is derived from the best estimates of *O*'s performance within each series, the mean of his matches. This correlation between means per *O* per series is .67, indicating a fair degree of accuracy of prediction of *O*'s score on Series B from his score on Series A. The reliability estimate of .67, however, indicates sources of variability not common to the tasks of the two series.

Summary

Windows of two sizes and two conditions for judgment were investigated for their effects on velocity matches. Two windows, one of which was twice the linear dimensions of the other, were used. Twenty-two *O*s viewed identical patterns through windows (the standard motion fields), and reproduced their velocities in terms of ft./sec. on the variable. Predictions were made about the expected differences between the means of the four conditions. Where no difference was predicted, no difference was significant. Five predicted differences were correct for direction, and three were significant ($P < .01$). The mean matches of all four conditions were significantly greater than the velocity of the standard. Accuracy of reproduction was greater in the small window of the variable than in the large window. The mean matches for the optimal and less favorable conditions were not significantly different while their variances were. The test-retest reliability of the matches for the less favorable conditions was significantly greater than for the optimal condition for reproducing.

References

1. Brown, J. F. The visual perception of velocity. *Psychol. Forsch.*, 1931, **14,** 199–232.
2. Cartwright, D. On visual speed. *Psychol. Forsch.*, 1938, **22,** 320–342.
3. Duncker, K. Uber induziente Bewegung. *Psychol. Forsch.*, 1929, **12,** 180–259.
4. Smith, O. W., & Gibson, J. J. Apparatus for the study of visual translatory motion. *Amer. J. Psychol.*, 1957, **70,** 291–294.
5. Wallach, H. On constancy of visual speed. *Psychol. Rev.*, 1939, **46,** 541–552.
6. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanimid Co., 1949. P. 14.

(Received May 20, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

THE DEPENDENCE OF PROBABILITY OF RESPONSE ON SIZE OF STEP INTERVAL IN THE METHOD OF LIMITS[1]

JACK BRACKMANN AND GEORGE COLLIER
Sacramento State College *University of Missouri*

Accounts of the visual threshold often assume that, under a fixed set of stimulating conditions, the probability of response to a simple, threshold, visual stimulus is a function only of the brightness and of random fluctuations in the intensity of the stimulus, or of the sensitivity of the organism, or both. If this assumption is correct, variations of the procedure of presenting the stimulus such as order, number, and pattern, should not influence the probability of response.

That this is not the case for the classical psychophysical methods is attested to by the many sorts of errors and biases which have been named, e.g., time error, error of habituation or expectation, error of central tendency, etc. The main import of these errors is that successive judgments are not independent. Some of these errors, e.g., the time error, have been intensively investigated. Others—for example, the errors of habituation and expectation—have been the subject only of admonitions and precautions.

Characteristically, accounts of these errors have been ad hoc attempts to explain unexpected deviations in threshold determinations. The question can be reversed, i.e.: What effects would a given mode or condition of stimulus presentation have on the proportion of "yeses" (P_Y), and the derived threshold? Most studies of this sort have been on the methods of absolute judgment (e.g. **11, 21**). More recently, studies examining the effect of number, order, distribution, intertrial interval of stimuli, etc., on other psychophysical methods have appeared (e.g., **1, 5, 6, 20**).

In particular Cotton and Verplanck (**5, 6, 20**) investigated the effects of the direction, length, intertrial interval, and method of termination of a series of stimulations on the parameters of the psychophysical function obtained by ML (method of limits). The present experiment explores the effect of a further procedural variable—namely, the size of the brightness step-interval between successive stimulations.

The interest in this variable was generated by the association between successive responses found in single-brightness modes of stimulus presentation (e.g., **19**). On the assumption that the amount of association was a decreasing function of difference in magnitude between successive stimuli, it was hypothesized that the so-called error of habituation in ML might result from such association and the size could be manipulated by manipulation of the step-interval.

METHOD

Apparatus.—The apparatus used presented binocularly a circular, 510 $\mu\mu$, patch, subtending

[1] This research was supported in part by ONR contract N6onr-180, T. O. IV, Project 143-253 at Indiana University, and in part by the Council on Research, Duke University. A portion of this paper was reported at the 1954 meetings of the American Psychological Association, New York.

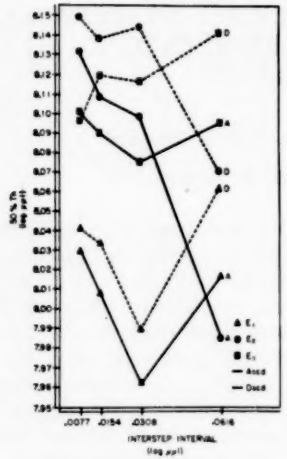

FIG. 1. The relation of the 50% threshold to the size of the brightness step-interval.

10′, 20° peripheral, for .001 sec., every 4.7 sec. The shutter operated continuously and provided no cue as to the occurrence of the flash. The range of brightnesses used was approximately 7–8 log $\mu\mu$l. A mouthbite was used. The *S* responded by the use of a key when he "saw" the flash. The apparatus was housed in a sound-deadened, light-tight, ventilated room. A more complete description of this apparatus can be found elsewhere (**2**).

Subjects.—Four trained *S*s were used. They were uniformly dark-adapted and were run at the same time each day for the 12 days of the experiment.

Procedure.—The ML procedure for presenting the stimuli consisted of alternating ascending (increasing brightnesses) and descending (decreasing brightnesses) series of brightness-ordered stimulations. Each brightness in a series, for a particular experimental condition, was separated by a fixed brightness increment, the step-interval. That is, the step-interval was the brightness increment or decrement between successive stimulations. Four step-intervals were used: .0077, .0154, .0308, and .0616 log $\mu\mu$l; each double the preceding value. The four *S*s were run over four days in a Latin-square design, each *S* receiving one of the four step-intervals on each of the days.

In order to control for the possible effects of total experimental time, number of determinations for each brightness, and for the length of the ascending and descending series, three procedures (E_1, E_2, and E_3) were used. In the first two the criterion for termination of an ascending series was three successive "yeses" in a row. The following descending series was started two step-intervals above the final "yes." A similar criterion was used for the descending series. In E_1 a total of 300 stimulations was given. In E_2, 15 runs of ascending and descending series were given. In E_1 and E_2 the range of brightnesses varied for each *S*. In E_3, two points were picked six step-units above and below the average threshold, determined from the first two experiments, and the series was terminated at these points. The range of brightnesses was held constant.

RESULTS

The proportion of "yeses" (P_Y) at each intensity for each session was plotted on probability paper and a straight line fitted by eye. The ascending series were considered separately. The 50% threshold and the 84%–50% brightness difference were interpolated from the fitted line. The latter value yields a measure which is inversely proportional to the slope of the fitted curve. This procedure is based on the ubiquitous assumption of a cumulative normal distribution of the P_Y. It yields the same results

TABLE 1

ANALYSIS OF VARIANCE OF 50% THRESHOLDS

Source	*df*	*MS*	*F*
Step-Interval (I)	3	.0049	.55
Ascending vs. Descending (D)	1	.0265	35.32**
Experiments (E)	2	.0791	4.38
Subjects (*S*)	3	.1349	
I × D	3	.0026	5.02*
I × E	6	.0116	1.60
I × *S*	9	.0090	1.24
D × E	2	.0008	2.53
D × *S*	3	.0008	1.6
E × *S*	6	.0180	2.46
I × D × *S*	9	.0005	.54
I × E × *S*	18	.0073	7.58**
I × D × E	6	.0016	1.62
D × *S* × E	6	.0003	.31
I × D × E × *S*	18	.0010	
Total	95		

* $P \leqq .05$.
** $P \leqq .01$.

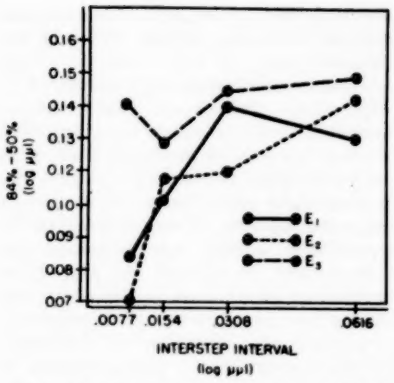

FIG. 2. The relation of the 84%–50% brightness difference to the size of the brightness step-interval.

as are obtained by cumulating the distribution of the points of alternation of response (e.g., from "yes" to "no") within the limits of the vagaries of graphic interpolation, the number of alternations in a series, and the assumption of normality.

50% threshold.—The results for the threshold measurements are presented in Fig. 1, in which thresholds are plotted against step-interval for each experiment and for ascending and descending series separately. An analysis of variance of step-interval, ascending vs. descending series, experiments, and Ss was performed (Table 1). In each experiment and for all step-intervals, with one exception that can be noted in Fig. 2, ascending series (from dimmer to brighter) yielded significantly ($P < .01$) lower thresholds than descending series. Further, the difference between the thresholds increases significantly ($P < .05$) with increasing size step-intervals. Experiment 1 yielded significantly ($P < .01$) lower thresholds than either Exp. 2 or 3. None of the other main effects or interactions were significant with the exception of the I × E × *S*.

84%–50% brightness difference.—The results for the measures of the 84%–50% difference are presented in Fig. 2 in which 84%–50% differences are plotted against step-interval for each experiment. An analysis of variance similar to the preceding was performed (Table 2). In each experiment and under all conditions the 84%–50% difference was an increasing function ($P = .05$) of the step-interval. That is, as the step-interval increases, the slope of the P_T function decreases; the judgments become less precise. The 84%–50% differences were significantly ($P = .05$) larger in E_3. None of the other main effects or interactions were significant with the exception of the I × E × *S*.

DISCUSSION

Three findings are of interest in the present study: the difference in ascending and descending thresholds, the decrease of this difference with a decrease in the interstep-interval, and the decrease in variability with a decrease of step-interval.

TABLE 2

ANALYSIS OF VARIANCE OF 84%–50% DIFFERENCES

Source	*df*	*MS*	*F*
Step-Interval (I)	3	.0092	4.85*
Ascending vs. Descending (D)	1	.0001	.10
Experiments (E)	2	.009	6.62*
Subjects (*S*)	3	.0032	
I × D	3	.0010	3.77
I × E	6	.0020	1.45
I × *S*	9	.0019	
D × E	2	.0003	1.47
D × *S*	3	.0011	
E × *S*	6	.0014	
I × D × *S*	9	.0003	.76
I × E × *S*	18	.0014	4.09**
I × D × E	6	.0006	1.76
D × *S* × E	6	.0002	.56
I × D × E × *S*	18	.0003	
Total	95		

* $P \leqq .05$.
** $P \leqq .01$.

These findings suggest that in a series of ordered stimulations two tendencies are present: a tendency to *alternate*, following a "long" sequence of responses of the same sort; and a tendency to *perseverate*— that is, to repeat the response given on the preceding trial(s).

Alternation.—The tendency to alternate in situations where similar or equivalent consequences follow the alternative responses or where the responses have been forced is well established (**15**). Of particular interest in the present analysis are the "guessing" experiments (**15, 16**) in which sequences of guesses have been shown to be nonrandom as a result of alternation, and the forced choice experiments (e.g. **7, 22**) which have shown that a tendency to alternate is generated by a sequence of forced choices of one of two alternatives. The former experiments are of interest since they show that tendency to alternate is independent of the occasion of the choice and dependent on the preceding response history. The latter experiments are interesting since they closely resemble psychophysical methods employing intensity-ordered sequences of stimulations. That is, any sequence of stimulations starting outside of the threshold region leads to a series of forced responses. This should, according to previous evidence, lead to a tendency to alternate irrespective of the stimuli presented on the subsequent trial(s).

The tendency to alternate in the ML situation was first parametrically studied by Cotton and Verplanck (**5, 6, 20**). They found a significant difference between ascending and descending series. The ascending series produced the lower thresholds. This difference in thresholds proved to be an increasing function of the length of series; that is, the earlier a series was initiated the sooner it terminated. These findings are fortified by the present study in which a highly significant difference, in the same direction, was again found between ascending and descending series. One major difference between their study and the present one was the size of the step-interval used, .1 log $\mu\mu$l. This value is greater than the largest value in the present study.

Reactive inhibition, stimulus satiation, exploratory tendencies, reinforcement consequencies, etc., have been proposed as explanations of alternation (**7, 8, 9, 12, 14, 15, 22**). Of these alternatives, the present data could conceivably support only the version of the reinforcement hypothesis which was proposed by Cotton and Verplanck (**5, 20**). This account of alternation assumes that the events associated with the termination of an ascending or descending run—that is, the appearance or disappearance of the stimulus—are reinforcers. Thus, the response "no" in a descending run and the response "yes" in an ascending run are reinforced. Further, the *sequence* of preceding "yeses" in a descending run and the sequence of preceding "noes" in an ascending run become descriminative stimuli for the association. The combination of these two factors would thus lead to a tendency to "anticipate"—that is, alternate.

Perseveration.—The second finding, the decrease of the difference between ascending and descending series with decreased step-interval, supports the original hypothesis of the study that the probability of making a given response ("yes" or "no") on a given trial is increased by the occurrence of that response on the preceding trial(s), and that this tendency is a decreasing function of the difference in magnitude between successive stimuli. Two recent studies (**3, 4**), employing a single-brightness mode of stimulation, have shown that this sort of intertrial association exists and is response-outcome-dependent rather than a result of fluctuations in sensitivity.

In the present study the tendency to perseverate is competing with the tendency to alternate. The outcome of this competition is reflected in the fact that the difference between ascending and descending thresholds decreases as the size of the step-interval decreases—that is, as the likelihood of the same response following a given response is

increased. This occurs even when the length of the series is greater for the small step-interval in E_2. This analysis is supported by a similar finding in the second study mentioned above (**4**), namely, that the probability of response on a given trial is a joint function of the magnitude of the stimulus, the preceding response, and the magnitude of the stimulus to which that response was made. As in the present study, the effect of the preceding response was greater, the greater the similarity in magnitude between the successive stimuli.

The interaction of these two tendencies can presumably be determined by manipulation of the length of series and the size of the step-interval. Thus, it should be possible to produce either errors of habituation or expectation and vary their magnitude by the proper combination of these parameters.

Variability.—The third finding, that of greater variability for the larger step-interval, also supports the original hypothesis of the study. The probability of a "yes" to a high brightness is high, as is the probability that it is preceded by a "yes." Thus, the joint probability of a "yes" to a given high brightness following a "yes" in an intensity-ordered series of stimulations is higher than that expected for a sample in which there is little or no association. Alternatively the probability of a "no" and the probability of its being preceded by a "no" is greater for a low brightness. Thus, if there is a greater association between successive responses to stimuli which are similar in magnitude, it would be expected that the probability of "yeses" at the high end and the probability of "noes" at the low end of an alternating ordered series of stimuli would be greater when the step interval was small. Thus, the series with a small step-interval would be expected to have a greater slope (smaller σ) as is observed in the present experiment. Any factor which increased the amount of intertrial association, such as a small intertrial interval (**3**), would be expected to increase the "precision" of judgment in an intensity-ordered series of stimulations.

The present analysis suggests that determinations of the psychophysical function employing methods having the greatest similarity of successive stimuli would have the lowest variability, i.e., greatest precision. Thus, the methods used, for example, by Hecht, Shlaer, and Pirenne (**10**) and Stevens, Morgan, and Volkmann (**17**) which employ blocks of stimuli of the same magnitude for the determination of each point would have less variability than those determined by method of limits, random methods, etc. Such a result has been obtained by Miller and Garner (**13**) in their comparison of the quantal method with the random method.

With the exception of two studies (**4, 18**), the previous findings on intertrial association between successive responses to threshold stimulation has been for data obtained by the single-brightness method. The present study suggests that it is possible to generalize these findings to situations in which the successive responses are to stimuli varying in magnitude. Two parameters governing the magnitude of the effect have been isolated; intertrial interval (**3**) and interstimulus difference. A further tendency was demonstrated in this and the Cotton-Verplanck studies (**5, 6, 20**), a tendency to alternate following a sequence of response to the same kind which appears to vary as a function of the length of the preceding series. It would seem most likely that an account of many sorts of "errors" observed in different psychophysical procedures could be given in terms of these phenomena.

Summary

The effect of the size of step-interval (size of brightness increment or decrement) on the parameters of the psychophysical function obtained by the method of limits was explored. Four highly trained *S*s were run through four absolute threshold determinations. One of four step-intervals was used for each determination. Three experiments using three variations of the ML were performed.

It was found that the threshold determined from ascending series (increasing brightness) were lower than those determined from descending series and that this difference was an increasing function of the step-interval. The variability of the psychophysical function was directly related to the size of the step-interval.

These results were interpreted as being the product of an interaction between a tendency for alternation, which is a function of the length of the preceding sequence of responses of the same kind, and a tendency for perseveration, which is a decreasing function of the size of the step-interval and of the preceding response. Evidence for the latter was reviewed from the single-brightness data. The relation of the variability to the size of the step-interval was attributed to the increase in probability of a "yes" (or "no") at the extremes of an intensity ordered series with a small step-interval due to the increase in intertrial association. The so-called errors of habituation and expectation were related to these findings.

References

1. Blackwell, H. R. Psychophysical thresholds: experimental studies of measurement. *Engr Res. Inst. Bull.*, Univer. of Michigan, 1952.
2. Collier, G. Probability of response and intertrial association as functions of monocular and binocular stimulation. *J. exp. Psychol.*, 1954, **47,** 75–83.
3. Collier, G. Intertrial association as a function of intertrial interval. *J. exp. Psychol.*, 1954, **48,** 330–334.
4. Collier, G., & Verplanck, W. S. Nonindependence of successive responses at the visual threshold as a function of interpolated stimuli. *J. exp. Psychol.*, 1958, **55,** 429–437.
5. Cotton, J. W., & Verplanck, W. S. The dependence of frequencies of seeing on procedural variables: II. Procedure of terminating series of intensity-ordered stimuli. *J. gen. Psychol.*, 1955, **53,** 49–57.
6. Cotton, J. W., & Verplanck, W. S. The dependencies of frequencies of seeing on procedural variables: III. The time-interval between successive stimuli. *J. gen. Psychol.*, 1955, **53,** 59–66.
7. Estes, W. K., & Schoeffler, M. S. Analysis of variables influencing alternation after forced trials. *J. comp. physiol. Psychol.*, 1955, **48,** 357–362.
8. Glanzer, M. Stimulus satiation: an explanation of spontaneous alternation and related phenomena. *Psychol. Rev.*, 1953, **60,** 257–268.
9. Heathers, G. L. The avoidance of repetitions of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10,** 359–380.
10. Hecht, S., Shlaer, S., & Pirenne, M. H. Energy, quanta, and vision. *J. gen. Physiol.* 1942, **25,** 819–840.
11. Helson, H. Perception. In H. Helson (Ed.), *Theoretical foundations of psychology.* New York: D. Van Nostrand, 1951, 349–389.
12. Miller, G. A., & Frick, F. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56,** 311–324.
13. Miller, G. A., & Garner, W. R. Effect of random presentation on the psychometric function: implications for a quantal theory of discrimination. *Amer. J. Psychol.*, 1944, **57,** 451–467.
14. Montgomery, K. C. A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, 1952, **45,** 287–293.
15. Solomon, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45,** 1–40.
16. Solomon, R. L. A note on the alternation of guesses. *J. exp. Psychol.*, 1949, **39,** 322–326.
17. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54,** 315–335.
18. Verplanck, W. S., & Blough, D. Randomized stimuli and the nonindependence of successive responses at the visual threshold. *J. gen. Psychol.*, in press.
19. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44,** 273–282.
20. Verplanck, W. S., & Cotton, J. W. The dependence of frequencies of seeing on procedural variables: I. Direction and length of series of intensity ordered stimuli. *J. gen. Psychol.*, 1955, **53,** 37–47.
21. Wever, E. G., & Zener, K. E. The method of absolute judgment in psychophysics. *Psychol. Rev.*, 1928, **35,** 466–493.
22. Zeaman, D., & House, B. J. The growth and decay of reactive inhibition as measured by alternation behavior. *J. exp. Psychol.*, 1951, **41,** 177–186.

(Received May 23, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

NONINDEPENDENCE OF SUCCESSIVE RESPONSES AT THE VISUAL THRESHOLD AS A FUNCTION OF INTERPOLATED STIMULI [1]

GEORGE COLLIER AND WILLIAM S. VERPLANCK [2]

University of Missouri *Indiana University*

The probability of "yeses" and "noes" to stimuli in the threshold region is some function of the stimulus magnitude. The variability of response to a stimulus of a given magnitude is typically attributed to random variables such as quantum fluctuations, fluctuations in neural sensitivity, etc. Recent studies of the randomness of response show that additional, nonrandom variables must be considered. Successive threshold responses to repeated, simple, visual stimuli of invariant brightness are not independent; "yeses" and "noes" occur in runs. The present experiment is concerned with the effect produced on this nonindependence by the interpolation of stimuli of higher or lower brightness.

Two kinds of intertrial association are possible (**3**). In the first the shifts in probability of response from trial to trial are some function of the response on the preceding trial or trials. In the second, the shifts are a function of some factor, e.g., excitability, varying in time, whose value is independent of the preceding response or responses. If the first kind of association obtains, the interpolation of stimulus brightnesses that yield high or low probabilities of "yes" into a series of constant-brightness trials, should alter the probability of "yes" on trials following these interpolations. If the second kind of association obtains, the interpolation of such response-forcing stimuli should not systematically alter the probabilities of response on subsequent trials.

Stimuli whose magnitudes fall well outside the range of a concurrent series of stimulations, such as the interpolated stimuli mentioned above, have often been invested with special properties. They are presumed to produce shifts in threshold, in the precision of judgments, or in both, by acting as *anchors*, by providing *frames of reference*, or by giving *S* a stable *criterion*.

Method

Apparatus.—The apparatus used was a Model III Hecht-Shlaer Adaptometer (**13**). This presented every 4 sec., binocularly, a 3° circular patch 7° below a red fixation point at a viewing distance of 25 cm. for a duration of .2 sec. The brightness of the patch was continuously variable from 1.5 log $\mu\mu$l. to 5.4 log $\mu\mu$l. The color of the patch was determined by a filter having a maximum transmittance at 480 mμ. Calibration was maintained photometrically. The flash duration was determined by a pendulum shutter, and the intertrial interval by a synchronous motor that operated the shutter. The apparatus was housed in a sound-deadened, light-tight, ventilated room.

Subjects.—Ten paid university undergraduates, 6 men and 4 women, served as *Ss*. Five of the *Ss* were from Indiana University, five from Duke University. All met or exceeded naval visual standards as tested by the Orthorator (20/20 visual acuity, O. D. O. S.; no phorias; "color" normal). All were previously inexperienced in visual experimentation.

The *Ss* remained uninformed of the purpose,

[1] This research was supported in part by ONR contract N6onr-180, T. O. IV, Project 143–253 at Indiana University, and in part by the Council on Research, Duke University. A portion of this paper was reported at the 1949 meetings of the Midwestern Psychological Association, Chicago, Illinois.

[2] Now at Stanford University.

procedure, and design of the experiment. They were instructed to maintain fixation and to say "yes" if they saw the flash, and "no" if they did not (Indiana *Ss*); or to push the lever to the right if they saw the flash, and to the left if they did not (Duke *Ss*). At the beginning of the experimental period each *S* had spent at least 20 min. in dark adaptation glasses and 15 min. in complete darkness. The *Ss* were run at approximately the same time each day.

Procedure.—Three procedures of stimulation were used. The first was a method of limits procedure (ML), consisting of 25 ascending and 25 descending brightness-ordered series of stimuli (**14**). The second was a single-brightness procedure, here termed the noninterpolation procedure (NI), consisting of two blocks of 75 stimulations of a single invariant brightness, selected from the ML results of each *S* to yield approximately 50% "yeses" ($P_Y = .50$). The third procedure, termed the interpolation procedure (I), was the same as the second, with the exception that after every five stimuli of the single brightness, a stimulus .5 log unit in brightness above or below the fixed value was presented. Fifteen high-brightness and 15 low-brightness stimuli were interpolated in a predetermined, semirandom order at the appropriate points in an experimental series of 150 fixed-brightness stimuli.

Experimental design.—Each *S* had three days of ML observations, during which the stimulus value to be used in the single-brightness procedure was determined. The procedure on each of the following six experimental days incorporated (*a*) five series of ML, (*b*) two blocks of 75 stimulations each, by one of the two single-brightness procedures, (*c*) two blocks of 75 stimulations each, by the other single-brightness procedure, and (*d*) five more series of ML. A 1–2 min. rest period intervened between each block of trials. The order of the two single-brightness procedures, I and NI, was alternated from day to day and was counterbalanced among *Ss*.

Results

The mean of the 50% thresholds determined for each *S* on the three preliminary days of ML was 2.50 log $\mu\mu$l. (SD = .15 log $\mu\mu$l.).

The average P_Y (proportion of "yeses") to the single-brightness stimuli over the six experimental days was .409 (s = .231) for the NI half of the experimental sessions, and .375 (s = .189) for the I half. These are significantly ($P < .01$) lower than 50%, indicating that in this experiment the brightness yielding 50% "yes" in the method of limits gives a lower P_Y when used in the single-brightness method. In previous results (**13**), closer agreement was found. They also differ significantly ($P < .05$), by a t test of related measures, from one another; the interpolation procedure slightly reduces P_Y to the single brightness.

Mode of analysis.—To examine the effects of the response to the interpolated stimuli and of the response preceding the interpolation, P_Y was determined for each of the five trials following each interpolation. Only those high interpolated stimuli to which a "yes" was given (mean P_{HY} = .975, s = .018), and those low interpolations to which a "no" was given (mean P_{LY} = .043, s = .026) were considered. The values of H and L were chosen so that such P_Y's would be obtained, i.e., they were chosen in the threshold range. These counts were first sorted into four groups, depending on what preceded them; the four cases considered were, (*a*) where a high interpolated stimulus was preceded by a "yes" ($_Y$H), (*b*) where a high was preceded by a "no" ($_N$H), (*c*) where a low was preceded by a "yes" ($_Y$L), and (*d*) where a low was preceded by a "no" ($_N$L). These values, averaged over days and *Ss*, are presented in the four curves of Part A of Fig. 1.

For purposes of comparison the same counts as above were performed on the data obtained under the NI procedure. Stimulations 6, 12, 18, etc., in the 150-trial blocks were treated as interpolated stimuli, and P_Y determined for each of the five trials following. A stimulus to which a "yes" was given was treated as a high interpolation, and a stimulus

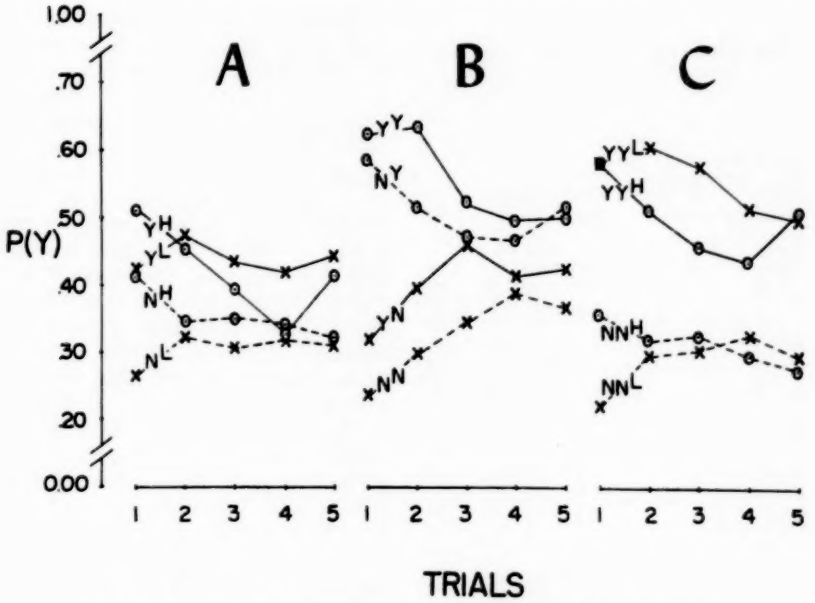

FIG. 1. Probability of response on trials following interpolation categorized for interpolation and response preceding.

to which a "no" was given was treated as a low. Hence, there are differing numbers of simulated H's and L's in the records. The same classification as before was made. These values, averaged over days and *S*s, are presented in the four curves of Part B of Fig. 1.

A second count was made on the data of the I portion of the records, in which the *two* responses preceding the interpolated stimuli were considered. The average P_Y for four of the eight cases, YYH, NNH, YYL, and NNL are presented in the four curves of Part C of Fig. 1. The averages for the other four cases, YNH, NYH, YNL, and NYL, deviated from the over-all average P_Y less than the comparable YY and NN curves, and are not presented.

The proportions graphed in Parts A and B of Fig. 1 were transformed to arcsin $\sqrt{p}$ and an analysis of variance was performed for the following variables: interpolated stimulus, high or low (I); response preceding an interpolation, "yes" or "no" (R); portion of the record, I or NI (P); trials following an interpolated stimulus (T); and *S*s (*S*). The analysis was made only on the first four of the five trials following interpolation, since the fifth trial was used to determine the response-preceding classification. The analysis is presented in Table 1. The main effects of I, R, and P were significant, as were the I × P, I × T, R × T, R × P × *S*, and P × T × *S* interactions. The I × R and I × P × T interactions approach significance. In both portions considered separately, the main effects of I and of R were significant.

Interpolation.—Consider first the simulated interpolations in the NI

TABLE 1

ANALYSIS OF VARIANCE OF THE ARCSINE P_Y ON TRIALS FOLLOWING INTERPOLATION

Source	*df*	*MS*	*F*
Interpolation (I)	1	3871.988	17.08**
Response preceding (R)	1	1782.461	6.84*
Portion of record (P)	1	2205.315	14.82**
Trials (T)	3	96.336	1.44
Subjects (*S*)	9	4584.566	
I × R	1	223.312	3.79
I × P	1	2299.118	25.99**
I × T	3	683.341	12.90**
I × *S*	9	226.736	3.56
R × P	1	19.474	.07
R × T	3	163.842	3.30*
R × *S*	9	260.734	.90
P × T	3	64.039	.55
P × *S*	9	148.771	.51
T × *S*	27	67.134	.58
I × R × P	1	19.831	.46
I × R × T	3	22.656	.53
I × R × *S*	9	58.910	1.38
I × P × T	3	91.099	2.13
I × P × *S*	9	88.451	2.07
I × T × *S*	27	52.947	1.24
R × P × T	3	21.257	.50
R × P × *S*	9	290.016	6.78**
R × T × S	27	49.605	1.16
P × T × *S*	9	116.582	2.72**
I × R × P × T	3	8.477	.19
I × R × P × *S*	9	65.510	1.47
I × R × T × *S*	27	33.586	.75
I × P × T × *S*	27	47.588	1.07
R × P × T × *S*	27	37.878	.85
I × R × P × T × *S*	27		
Total	301		
Pooled error	117	42.802	

* $P \leqq .05$.
** $P \leqq .01$.

portion of the records, without regard to the response preceding them: When P_Y is sampled following a response to a stimulus in a single brightness series, P_Y is significantly *greater* than expected following a "yes," and it is significantly *less* than expected following a "no" (B, Fig. 1). The divergence from the expected value becomes less over the succeeding trials following the simulated interpolation (I × T interaction). Successive responses are not independent, as has often been pointed out.

Consider now the interpolation (I) portion of the record: When P_Y is sampled following a response to an interpolated stimulus of different brightness in a single brightness series, P_Y is significantly *greater* than expected following a "yes" response to a brighter stimulus, and it is significantly *less* than expected following a "no" to a dimmer one (A, Fig. 1). That is, the "yeses" and "noes" forced by the interpolated stimuli shift the values of P_Y in the expected direction. Long series of "noes" appear to be more affected than long series of "yeses" (C, Fig. 1).

The divergence from the expected value of the P_Y following the interpolated stimuli is significantly less (see I × P interaction) than that for P_Y's following the simulated interpolations. On the first trial following an interpolation in the NI portions of the record, the P_Y following a "yes" is .63 and following a "no" is .27, while in the I portions the P_Y following a "yes" (to an H) is .47 and following a "no" (to an L) is .33. The differences between the P_Y's are significant for both the "yes"-H and "no"-L comparison. The differences, .19 and .06, are themselves significantly different. Thus, the probability of a "yes" following a "yes" to a stimulus of the same brightness is greater than the probability of a "yes" following a "yes" to a stimulus of a different (brighter) brightness. The probability of a "no" following a "no" to a stimulus of the same brightness is greater than the probability of a "no" following a "no" to a different (less bright) brightness, but the difference is significantly less than that for the "yeses."

The significant decline to the expected value (I × T interaction) is more rapid for the P_Y's preceded by

the interpolated stimuli than for those preceded by the simulated interpolated stimuli (see I × P × T interaction and Fig. 1).

Response preceding.—A "yes" preceding an interpolation is associated with a greater than average P_Y on the trial subsequent to the interpolation, a "no" preceding an interpolation is associated with a smaller than average P_Y on the trial subsequent. The effect is less than that produced by the interpolated stimulus and declines slowly over three or four trials to the expected value (R × T interaction and Fig. 1). The differences between the ${}_Y$H and ${}_N$H, the ${}_Y$L and ${}_N$L, and the ${}_Y$Y and ${}_N$Y curves are all significant at at least the .05 level using the appropriate error term from the analysis of variance. The failure of the R × P interaction to approach significance reflects the expected finding that the response preceding an interpolation, real or simulated, has the same effect on the response subsequent to the interpolation in both I and NI procedures.

The R × I interaction approaches significance: The differences between high and low interpolations are *less* when preceded by a "yes" than when preceded by a "no"; the effect may not be symmetrical. This rather surprising finding has implications for psychophysical methodology and needs to be investigated further. Saying "no" is rather different from not saying "yes."

The average P_Y's for two responses preceding the interpolations in the I portion of the records are presented in Part C of Fig. 1. Since they are subsamples of the curves in Part A of Fig. 1, they resemble these curves very closely, with the exception that they are all significantly displaced farther away from the expected values. It is again apparent from these curves that the association between responses to stimuli of fixed brightness is larger than that of responses to stimuli of higher or lower brightness inserted in such a series.

The interpolation of stimuli of higher and lower brightnesses has two further effects. It results in a lower over-all P_Y, and it results in a significantly reduced, $P < .001$ ($F = 15.83$; $df = 1, 59$), amount of intertrial association as measured by a test based on the number of runs (**9**).[3] That is, the interpolated stimuli tend to break up the "runs" of responses. An omnibus χ^2 test (**2**) shows that the amount of association in the I portion of the record, while reduced, is still present at at least the .01 level.

Discussion

The fact that a "yes" response is more likely to follow a "yes" and a "no," a "no" as is shown in the NI portion of the data can be accounted for on either of two assumptions (**3**): that P_Y is being sampled from a population in which momentary shifts in sensitivity[4] producing nonrandom runs of responses are occurring independently of *S*'s pre-

[3] The response to the interpolated trials was left out in calculating the number of runs. This may result in an error due to the different intertrial intervals. By comparison with earlier data (**2**), the reduction could not be accounted for by the increased intertrial interval. The autocorrelation formulae based on the coefficient is related to the critical ratio formulae $N\phi^2 = \chi^2 = CR^2$. This applies to the approximation of the Wald-Wolffitz runs test used by Collier (**3**) and McGill (**9**) with the goodness of the assumption that R is distributed normally, and exactly for the test used by Collier (**2**).

[4] In order for the contingent probabilities to be as high as they are in the present experiment, the shifts in sensitivity about the obtained mean value would have to be very large, i.e., shifts in P_Y up to at least .80 and down to at least .15.

ceding responses, or that P_Y on any trial is controlled by some feature of the immediately preceding response(s). However, if a "yes" or a "no" is obtained, not by sampling but by "forcing"—that is, by the presentation of a stimulus of a magnitude such that a "yes" or a "no" is highly probable—then the hypothesis of shifts in sensitivity would not predict that the P_Y following the forced "yes" or "no" would differ from the average, since, presumably, both kinds of interpolated stimuli would have occurred equally often in periods of both high and low sensitivity. Forced "yeses" and "noes" alter P_Y. The response-linkages observed are some function of the immediately preceding response or responses. The same conclusion was reached in a previous study by Collier (**3**), in which he showed that the form of the function relating the amount of intertrial association and the intertrial interval was that expected on the response-outcome hypothesis. However, the results of the present study show that a third kind of linkage must be considered.

Intertrial association is not solely a function of the partial determination of a response by the preceding response or responses; the difference between the P_Y's following interpolations and those following simulated interpolations indicate that this contingency is a joint function of both the preceding response and the stimulus to which it was made. As Parts A and B of Fig. 1 show, for both "yeses" and "noes," the amount of linkage is greatest when the stimulus intensity is invariant. The probabilities of response determined by single-brightness methods are not equivalent to those obtained when stimulus intensity is systematically varied. Response-to-response linkage is maximized in single-stimulus methods, by the procedural imposition of invariant stimulation. This is not to say that such linkages do not exist in other methods (cf. **1, 13**), but merely that they are not maximal.

This greater response-response linkage is reflected in the smaller variability in threshold determinations by the single-brightness methods.[5] This implies that the methods, for example, used by Hecht, Shlaer, and Pirenne (**8**) in determining the absolute threshold, in which stimuli are presented in blocks or the period of observation is "long," would result in lower variability. The arguments based on the lower variability obtained, i.e., the quantum hypothesis of Stevens, Morgan, and Volkmann (**12**) (cf. Footnote 5) or the determination of the limits of biological variability by Hecht, Shlaer, and Pirenne (**8**) or Crawford and Pirenne (**4**) would be vitiated by this interpretation. This interpretation does not imply that the increase in response linkage is an increase in error, but rather may be an increase in sensitivity. This point will be discussed later.

A number of different studies support the hypothesis that intertrial association is a *joint* function of the response preceding and the stimulus to which it was made. Verplanck and Blough (**13**) analyzed the contingencies in visual threshold data obtained by the constant method in an experiment where stimulus brightness was systematically varied from trial to trial, in terms of both the stimulus and of the response preceding. They found the typical dependencies for response preceding and no relation for the stimulus preceding, but, when the data were further categorized for

[5] Hecht, Shlaer, and Pirenne (**8**), and Stevens, Morgan, and Volkmann (**12**), and others, have used variations of the single stimulus method, typically presenting their stimuli in blocks all of the same magnitude. The present results would predict that such methods would show lower variability as they suggest. This reduction in variability is the result of the maximal response-response association resulting from the use of (blocks of) invariant successive stimuli rather than the establishment of a "criterion," etc., as has been suggested. Experimental support for this predicted reduction in variability has been reported by Miller and Garner (**11**) in their comparison of the "quantum" and "random" method of stimulus presentation. They interpret the increase in variability in the quantum method in terms of the assumption of a "three quanta criterion" rather than a two for the random method.

the stimulus intensity to which a "no" was given on the preceding stimulation, they did find a relation. A "no" to a high intensity stimulus resulted in a lower P_Y on the subsequent trial. This is similar to the present findings for L interpolation vs. N. Garner (**5**) reports similar findings for absolute loudness judgments, in which information transmitted from trial to trial (which reflects nonindependence) increases both as a function of the preceding stimulus and of the preceding response, not considered jointly. The size of the increase was an increasing function of the number of stimulus categories used. According to Garner, at least a portion of the increase in transmission is the result of the joint action of the stimulus and the response preceding; further analysis was not practical because of its complexity. McGill (**9**), using a forced-choice situation with four masked tones, found that 28% of the response information was due to association between *S*'s response, the stimulus to which it was made, and the preceding response; of which half (14%) was contributed by the interaction between the stimuli and the preceding response. Here the relationship between the current stimulus and the preceding response is evaluated without respect to the stimulus to which the preceding response was made. In a recent analysis (**10**) of the same situation he reports that amount of information transmitted by the preceding response was a decreasing function of the frequency with which stimuli were correctly judged. Goldiamond (**6**) in a forced two-choice situation reports a relation between the preceding stimulus intensity and both the probability of making a given choice and the confidence of having seen the current stimulus (but not the probability of correctly locating the stimulus).

These findings may be interpreted as indicating that responses to stimuli near the threshold are determined less by the stimulus and more by the preceding response(s). This interpretation is made by McGill (**10**) in the study mentioned above. But, according to the present findings, stimulus similarity may be a more important variable, and the amount of association may be relatively independent of the general level of intensity. That is, the more similar successive stimuli are to one another in magnitude or quality [cf. Garner (**5**)], the greater the association between successive responses, even though *S* is not required to discriminate between them.

When the response-outcome model is accepted one may ask how many preceding responses have to be considered in order to account for the association found in the single-stimulus methods. Weiss, Coleman, and Green (**15**), on the basis of their results and from considerations of parsimony, proposed a "first order" auto-regressive model for data obtained from successive bisections of an angle. The complexity of information theory analysis has led others as well to consider only the first response preceding. Our analysis on the basis of the two responses preceding interpolation require consideration of *at least* the second response preceding *as well as* the magnitude of the stimuli to which the preceding responses are made. A one-step model is inadequate: while the difference between ${}_Y$Y and ${}_N$Y and between ${}_Y$N and ${}_N$N can be accounted for on sampling grounds, the difference between the ${}_Y$H and ${}_N$H and between ${}_Y$L and ${}_N$L cannot. This is consistent with results (**3**) on the relation between the intertrial interval and the amount of association at various lags; at least the second preceding response had to be considered in order to account for the differences between the curve relating amount of association and intertrial interval for various lags.

Few attempts at explanation of intertrial association have been made since the physiological and physical ones reviewed by Guilford (**7**) with the exceptions, perhaps, of McGill's (**10**) characterization of them as guessing habits,[6]

[6] This is an unfortunate use of terminology since the nonrandomness observed in experiments on guessing has typically resulted from alternation rather than perseveration (**1**).

or the stochastic model developed by Weiss, Coleman, and Green (**15**). The latter model suffers from so few empirical constraints that it cannot be considered useful as an explanation.

Guilford's study (**7**) and that of Collier (**2**) would seem to eliminate any simple physiological or physical explanations as, for example, eye movements, fluctuations in pupil size, etc. The low intercorrelations between modalities obtained by Wertheimer (**16**) suggest that there may be some basic underlying physiological factor related to the dependencies, but the small amount of variance accounted for would make it almost negligible. The results of the present study, while not precluding an account in terms of sensitivity drift, would imply that it is of small magnitude and acts jointly with the response-response and stimulus-response-response dependencies. Verplanck and Blough (**13**) interpret their data as supporting the former type dependencies but not unequivocally. Weiss, Coleman, and Green (**15**) find some evidence for a slow drift using power-spectrum statistics. In any case, the present results leave this an open question.

Three and possibly four empirical constraints can be placed on any future model. The magnitude of association (*a*) is related to the similarity between successive stimuli, (*b*) is related to the general level of stimulus intensities employed (**10**), (*c*) is a decreasing function of the intertrial interval (**3**), and (*d*) may be related to the number of categories of judgment (**5**). Whether *a* is really a special case of *b* remains an experimental question. One explicit test of the hypothesis of the relation between response-association and stimulus similarity appears to confirm it (**1**). On the assumption that the tendency to respond in the same fashion is related to the stimulus difference, it was expected that in method of limits employing different brightness step intervals, the point of termination in both ascending and descending series would occur later for the series with the smallest step interval. Such proved to be the case. The studies of Garner (**5**) and McGill (**9, 10**) also seem to support the "similarity" hypothesis.

One may ask whether the runs of responses reflect an increased accuracy of reporting, or the opposite: an increase in "false" positives and negatives. The stimulus-similarity hypothesis discussed above, contrary to the usual notions, suggests that although runs of "yeses" reflect increased accuracy, runs of "noes" in part reflect a decrease; that is, stimuli that were "seen," not being reported (cf. the difference between L—interpolation and simulated L—interpolations). Goldiamond (**6**) reports that sequential dependencies do not lead to increased accuracy of location. In the present study the reduction of the over-all P_Y in the I portion of the record may result from either a reduction in *false* positives or from an increase in the number of stimuli "missed." The fact that there were very few "yes" responses to the L interpolations supports the latter interpretation.

Three of the findings reported here—the effect of the interpolated stimuli, the reduction of P_Y, and the over-all reduction of the sequential effect in the interpolation portions of the records—are suggestive in regard to the so-called "anchors," "frames of reference," or "criteria." Stimuli occurring well outside the range of the concurrent stimuli may produce shifts in response probability through modification of response-association, through reduction in the over-all nonindependence by breaking up runs of responses, and possibly a change in the proportion of false positives and false negative responses.

Summary

Stimuli of high and low brightnesses were interpolated in a series of invariant brightnesses in an experiment on intertrial association. The probability of a "yes" to a stimulus following a high interpolation is increased and following a low, decreased. These contingencies were compared with those obtained when certain of the stimuli in a series of invariant brightnesses were analytically treated as interpolated stimuli. The magnitude of intertrial association was

greater with the simulated interpolations than with real ones. That is, interpolated stimuli reduced the sequential effect. It is concluded that the probability of a given response to a stimulus is a joint function of the magnitude of the current stimulus, the response to the preceding stimulus, and the (relative) magnitude of the preceding stimulus. It is argued that the response dependencies extend beyond one trial, and they are probably not attributable to sensitivity fluctuations. Some of the empirical constraints on an explanation of intertrial association are discussed.

References

1. Brackmann, J., & Collier, G. The dependence of probability of response on size of step interval in the method of limits. *J. exp. Psychol.*, 1958, **55,** 423–428.
2. Collier, G. Probability of response and intertrial association as functions of monocular and binocular stimulation. *J. exp. Psychol.*, 1954, **47,** 75–83.
3. Collier, G. Intertrial association as a function of intertrial interval. *J. exp. Psychol.*, 1954, **48,** 330–334.
4. Crawford, G. H., & Pirenne, M. H. Steep frequency of seeing curves. *J. Physiol.*, 1954, **126,** 404–411.
5. Garner, W. R. An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, 1953, **46,** 373–380.
6. Goldiamond, I. Serial effects as a function of the type of indicator used. Paper read at Midwest. Psychol. Ass., Chicago, 1955.
7. Guilford, J. P. Fluctuations of attention with weak visual stimuli. *J. gen. Psychol.*, 1926, **38,** 534–583.
8. Hecht, S., Shlaer, S., & Pirenne, M. H. Energy, quanta, and vision. *J. gen. Physiol.*, 1942, **25,** 819–840.
9. McGill, W. J. Multivariate information analysis. *Psychometrika*, 1954, **19,** 97–116.
10. McGill, W. J. Serial effects in auditory threshold judgments. *J. exp. Psychol.*, 1957, **53,** 297–303.
11. Miller, G. A., & Garner, W. R. Effect of random presentation on the psychometric function: implications for a quantal theory of discrimination. *Amer. J. Psychol.*, 1944, **57,** 451–467.
12. Stevens, S. S., Morgan, C. T., & Volkmann, J. Theory of the neural quantum in the discrimination of loudness and pitch. *Amer. J. Psychol.*, 1941, **54,** 315–335.
13. Verplanck, W. S., & Blough, D. Randomized stimuli and the non-independence of successive responses at the visual threshold. *J. gen. Psychol.*, in press.
14. Verplanck, W. S., Collier, G. H., & Cotton, J. W. Nonindependence of successive responses in measurements of the visual threshold. *J. exp. Psychol.*, 1952, **44,** 273–282.
15. Weiss, B., Coleman, P. D., & Green, R. F. A stochastic model for time-ordered dependencies in continuous scale repetitive judgments. *J. exp. Psychol.*, 1955, **50,** 237–244.
16. Wertheimer, M. An investigation of "randomness" of threshold measurements. *J. exp. Psychol.*, 1953, **45,** 294–303.

(Received May 23, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

DIMENSIONAL ANALYSIS OF MOVEMENT REACTIONS[1]

EDWIN A. FLEISHMAN

Yale University

In previous studies the interrelationships among performances on a wide range of psychomotor tasks have been investigated (**2, 6, 8, 11, 12**). A primary concern has been to provide a functional classification of abilities accounting for individual differences in such skills.

The present study is concerned with a class of psychomotor skills which is probably the most important and contains the most numerous and complex kinds of human movements. These fall under the general heading of *movement reactions*, as defined earlier by Brown and Jenkins (**1**). In movement reactions one is interested in such things as the ability to make smooth or coordinated control movements, to move a body member or control at a given rate, in a rhythmical fashion, in a certain sequence, or along specified pathways. The distinguishing feature is that skill *during* the movement is of primary interest, as contrasted with *positioning movements*, where terminal accuracy is the primary feature, and *static reactions*, where maintenance of a certain limb position is the central task. Thus, although the act of reaching out to a given position in space to a control lever (positioning reaction) is the first step in operating that lever, the manner in which it is moved is the more significant feature of the reaction.

The classification of motor abilities into static, positioning, and movement reactions by Brown and Jenkins (**1**) is entirely a rational classification. Each of these areas, in turn, they subdivided into more restricted motor-response categories. A recent study (**5**) has investigated the utility of the rational distinctions made among static and positioning reactions through an analysis of intercorrelations among performances on representative tasks of these skills. The general finding was that from the point of view of individual differences, static reactions are usefully considered as involving different abilities from positioning reactions, but that different positioning tasks involve highly specific skills with little generality from one task to another. On the other hand, in another study involving movement reactions (**11**), broad group factors were found which contributed to skill on many superficially diverse kinds of tasks.

These group factors, empirically derived from the intercorrelations among tasks, did not conform in any simple manner to the subcategories of movement reactions outlined by Brown and Jenkins (**1**). For example, movement reactions are categorized by them into discrete, repetitive, serial, and continuous types. Our results indicated that performance on tasks representing these "subcategories" grouped into somewhat different kinds of categories which cut across those provided by Brown and Jenkins.

[1] This research was performed while the writer was with the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas, in support of Project 7707. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the U. S. Government. The writer gratefully acknowledges the assistance of Walter E. Hempel, Jr., in the conduct of this study.

The present study is a more definitive follow-up of the previous study involving movement reactions. The earlier study was based on the intercorrelations among performances on 24 psychomotor tasks, originally administered to over 1,000 Navy pilot candidates in 1947. In the present study, tasks were specifically selected or designed around certain hypotheses derived from the earlier study. These tasks were administered in the laboratory, under controlled conditions, to an unselected sample of basic trainee airmen in 1954.

The objectives of this study were: (*a*) to replicate the earlier study under more controlled conditions, (*b*) to obtain more precise definitions of the factors identified through the inclusion of certain additional tasks, and (*c*) to investigate the relationships between the factors identified and performance on a complex task performed under different conditions of difficulty.

Hypothesized Factors

The tasks were selected or constructed with a view to certain of the factors identified in the previous analysis of complex psychomotor tasks (**11**). These factors and their tentative definitions from this previous study are as follows:

1. *Fine Control Sensitivity.*—The ability to make fine, highly controlled limb adjustments in movements of moderate scope (labeled Psychomotor Coordination I in previous study [**11**]).

2. *Multiple Limb Coordination.*—The ability to coordinate more gross movements, where the use of more than one body member simultaneously is required (labeled Psychomotor Coordination II in previous study [**11**]).

3. *Rate Control.*—The ability to make continuous anticipatory motor adjustments relative to changes in speed and direction of a continuously moving target or object.

4. *Response Orientation.*—The ability to make rapid discriminations as to direction and choice of movements.

On the basis of the previous analysis (**11**), at least three tasks were included to sample each of these ability categories. Certain tasks in the previous analysis as well as new ones were included in order to clarify the generality of these factors and to sharpen the original definitions. For example, tasks thought to represent the *Fine Control Sensitivity* factor were, Rotary Pursuit, Complex Coordination, and Rudder Control, Pursuit Confusion, and Two-Hand Coordination. The task called Control Adjustment was a new device constructed with the definition of the Fine Control Sensitivity factor in mind and will be described later. The Dial Setting test was included, since this seemed to require highly controlled wrist-finger movements and its possible loading (or lack of loading) on Fine Control Sensitivity would help determine the generality of this factor.

Three tasks which measured *Multiple Limb Coordination* in the previous study were Plane Control, Two-Hand Coordination, and Complex Coordination. The Two-Hand Pursuit task, and the two Rudder Control scores were added in the present study as possible measures of this factor.

Tasks assumed to measure the *Rate Control* factor were the Single Dimension Pursuitmeter, Two-Hand Pursuit, and Rate Control tasks. The task called Motor Judgment was added since it appeared to involve a greater degree of rate judgment and anticipation than other tests of this factor. If our original definition of this factor was correct, then this test should be among the better measures of this factor. The Visual Coincidence test, which requires only a button-pressing response, was included to see if emphasis in this factor is mainly in judgment of rate of the stimulus rather than the response.

The *Response Orientation* factor was thought to be represented by the Discrimination Reaction Time, Printed Discrimination Reaction Time, and Dial Setting tests.

Tasks representative of three other factors were also included in the study. Tasks in the present study, which have loadings on these factors in previous studies in our series, are indicated in parentheses.

5. *Reaction Time.*—The speed with which an individual is able to respond to a stimulus when it appears (see **2, 10**). (Visual, Auditory, Jump Visual, and Jump Auditory Reaction Time tests.)

6. *Arm-Hand Steadiness.*—The extent to which an individual is able to make steady, restricted arm-hand movements of the type which minimize strength and speed (**2, 6, 7**). (Track-Tracing and Steadiness-Precision tests.)

7. *Speed of Arm Movement.*—The speed with which gross arm movements, of limited scope, can be made (**2, 9, 10**). (Rotary Aiming, Jump

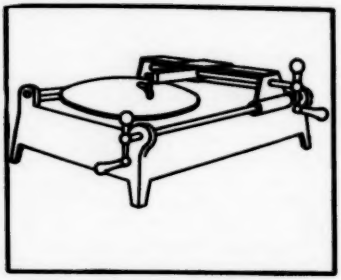

FIG. 1 TWO HAND COORDINATION

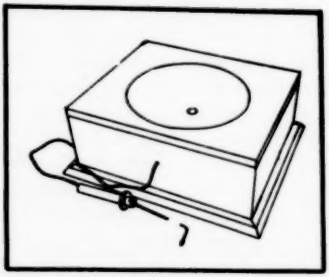

FIG. 2 ROTARY PURSUIT

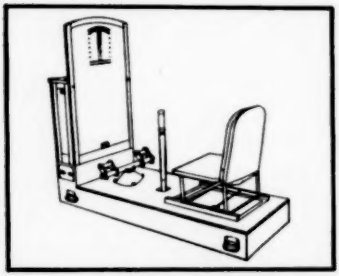

FIG. 3 COMPLEX COORDINATION

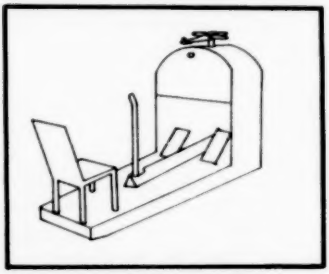

FIG. 4 PLANE CONTROL

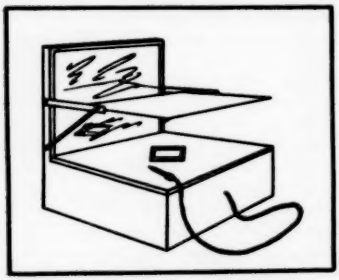

FIG. 5 PURSUIT CONFUSION

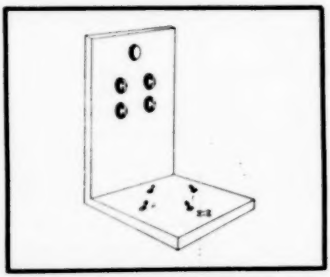

FIG. 6 DISCRIMINATION REACTION TIME

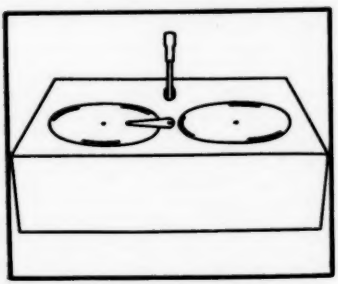

FIG. 7 MOTOR JUDGMENT

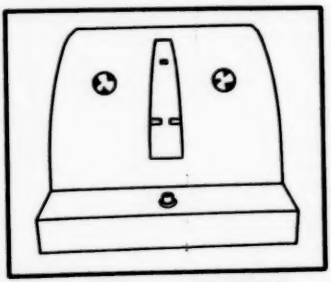

FIG. 8 VISUAL COINCIDENCE

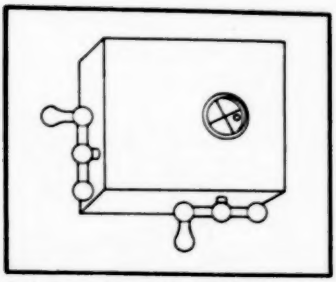

FIG. 9 TWO HAND PURSUIT

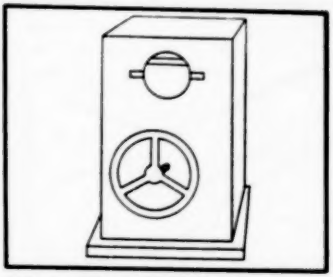

FIG. 10 SINGLE DIMENSIONAL PURSUIT

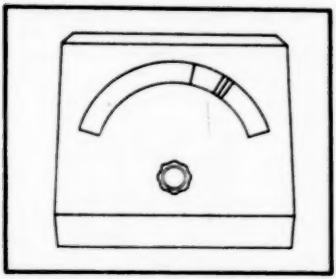

FIG. 11 RATE CONTROL

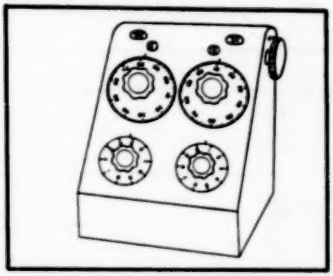

FIG. 12 DIAL SETTING

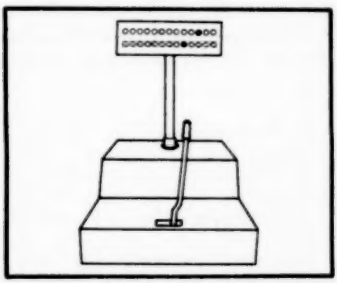

FIG. 13 CONTROLS ADJUSTMENT

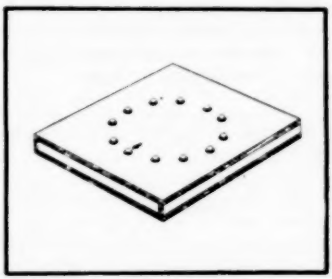

FIG. 14 ROTARY AIMING

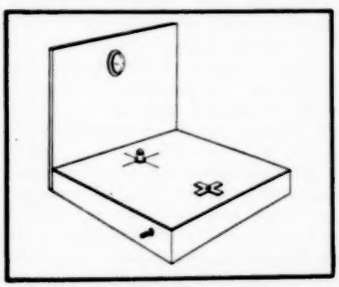

FIG. 15 REACTION TIME

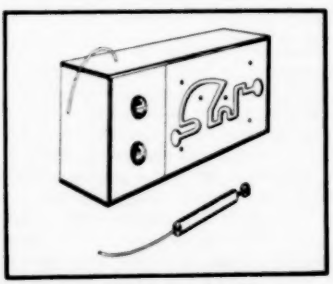

FIG. 16 TRACK TRACING

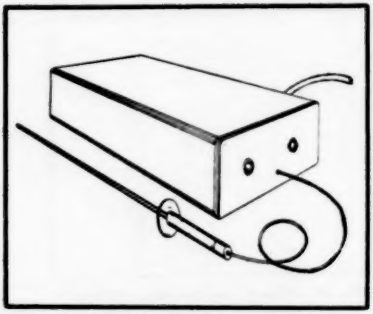

FIG. 17 STEADINESS PRECISION

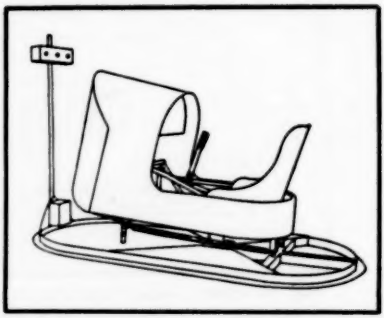

FIG. 18 RUDDER CONTROL

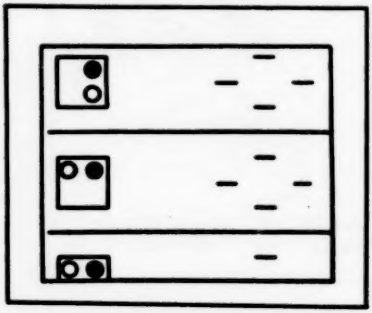

FIG. 19 PRINTED DISCRIMINATION REACTION TIME

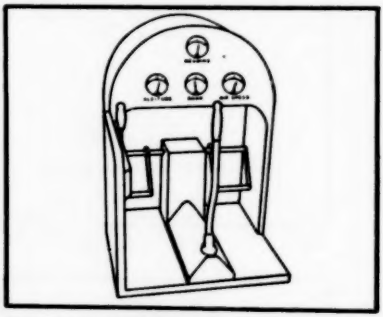

FIG. 20 MULTI-DIMENSIONAL PURSUIT

Visual Reaction Time, and Jump Auditory Reaction Time tests.)

Method

Description of the Experimental Tasks

Brief descriptions of the 31 task variables follow. References to more detailed descriptions are given where available:

1. *Two-Hand Coordination* (**13**) (Fig. 1).—The *S* attempts to keep a target-follower on a small target disk as the target moves irregularly and at varying rates. In the present study the target was reduced to .5 in. diameter. Movement of the target-follower to the right and left is controlled by one lathe-type handle; movement to and from is controlled by the other handle. Consequently, simultaneous rotation of both handles moves the follower in any resultant direction. Score was the total time on target during four 1-min. trials separated by 15-sec. rests.

2. *Rotary Pursuit* (**12**) (Fig. 2).—The *S* attempts to keep a prod-stylus in contact with a small metallic target, set in a rapidly revolving phonograph-type disk. Score was the total time on target during five 20-sec. trials separated by 10-sec. rests.

3. *Complex Coordination* (**13**) (Fig. 3).—Patterns of lights are presented whose positions are to be matched by appropriate adjustments of stick and rudder controls. A correct response is accomplished only when both the hands and feet have completed and maintained the appropriate adjustments, at which point a new pattern of lights to be matched is presented. Score is the number of completed matchings during two 2-min. test periods separated by a 30-sec. rest.

4. *Plane Control* (**14**) (Fig. 4).—The attitude of a model airplane is varied irregularly in its roll, pitch, and yaw axes by a motor-driven cam system. The *S* attempts to keep the airplane in a straight-and-level attitude by making compensatory adjustments of stick and pedal controls. Score is the amount of

time *S* keeps the plane straight and level (on target in all three components) during four 1-min. test periods, separated by 15-sec. rests.

5. *Pursuit Confusion (Time on Target)* (**11**) (Fig. 5).—The *S* attempts to keep a stylus on a variable-speed target as it moves through a diamond-shaped slot. The task is complicated by the fact that the entire target area is visible only by mirror-vision. Score is the time on target during four 1-min. test periods separated by 15-sec. rests.

6. *Pursuit Confusion (Errors).*—Same as above, but score is the amount of time *S* is in contact with the sides of the slot.

7. *Discrimination Reaction Time* (**12**) (Fig. 6).—The *S* manipulates one of four toggle switches as quickly as possible in response to a series of visual stimulus patterns differing from one another with respect to the spatial arrangement of their component parts (e.g., position of a lighted red lamp relative to a lighted green lamp). Score is the cumulated time of response for a series of 40 reactions for each stimulus pattern (*S* must respond within 3 sec.). Each set of 20 reactions was separated by a 35-sec. rest.

8. *Motor Judgment* (**12**) (Fig. 7).—The *S* is confronted by two adjacent disks rotating at a constant speed. Each disk has black and white sections on its perimeter. Between these disks is a pointer, whose speed of rotation *S* can control. A forward movement of this stick slows the pointer and a backward movement speeds up the pointer. The *S* cannot stop the rotation of this pointer completely and he can exert no control over the two rotating disks. The *S* is required to manipulate the control stick so as to make as many revolutions of the pointer as possible without crossing the black areas on the rotating disks. To do this properly he must integrate his estimates of the speed of each disk, the pointer, and his own control movements. Score is the ratio of number of pointer revolutions to number of errors (crossings of black areas) during four 1-min. trials separated by 15-sec. rests.

9. *Visual Coincidence* (**13**) (Fig. 8).—The *S* must react to a thin bar of light which moves from top to bottom behind a curved, translucent vertical screen. When the position of the moving light passes exactly between a pair of adjacent, stationary lights, *S* responds by pushing a button. If the reaction of *S* occurs at the exact point of coincidence, two red signal lamps at the sides of the screen flash. If *S* presses the button too soon or too late, no correct response is recorded (even if *S* holds the button down). For each of 16 stimulus presentations, the reference lights may appear at different places on the screen and the bar of light may move at different rates. Score is the number of correct responses for 64 stimulus presentations, with a 10-sec. rest between each series of 16 presentations. Each series of 16 presentations comprised 100 sec.

10. *Two-Hand Pursuit* (**12**) (Fig. 9).—This is similar in its response aspects to Variable 1, except this task involves compensatory rather than following pursuit. The *S* is required to coordinate the movements of two control handles to keep a small target under a cross-hair as it deviates from a null position. Score is the time the target is kept "on target" during four 1-min. trials separated by 15-sec. rests.

11. *Single Dimension Pursuitmeter* (**12**) (Fig. 10).—The *S* makes compensatory adjustments (in and out movements) of a control wheel, in order to keep a horizontal line in a null position as it deviates from center in irregular fashion. The control wheel is dampened pneumatically, introducing a lag into the system. Score is the time the horizontal line is held in a null position during four 1-min. trials separated by 15-sec. rests.

12. *Rate Control* (**12**) (Fig. 11).—A vertical target line moves back and forth across a curved scale, with frequent changes in direction and rate of movement. The *S* attempts to keep a thin pointer in coincidence with this line by adjustive rotary manipulations of a knob control. Score is total time the pointer and target line are in coincidence during four 1-min. trials, separated by 15-sec. rests.

13. *Dial Setting* (**12**) (Fig 12).—The *S* is required to set four dials to the exact numbered positions indicated in four stimulus apertures, one of which corresponds to each dial. When all four dials are set exactly according to the indicated numbers, four new stimulus numbers appear in the apertures. Score is the number of settings completed in three 2-min. trials, separated by 15-sec. rests.

14. *Control Adjustment* (Fig. 13).—The *S* is required to match the position of a red light with a corresponding green light. The position of the lighted green light is controlled by a highly sensitive control stick. A slight movement of the stick to the right displaces the lighted green light to the right and a slight movement of the stick to the left displaces the green light to the left. When *S* has matched the two lights and held this position for .5 sec., the red light moves to a new position and *S* proceeds to match it. Score is the number of completed matches in four 1-min. trials separated by 20-sec. rests.

15. *Rotary Aiming* (**2**) (Fig. 14).—The task is to strike at a series of buttons arranged in a circular pattern on a horizontal panel going from one button to the next as rapidly as pos-

sible. Score is the number of strikes in four 30-sec. trials separated by 25-sec. rests.

16. *Visual Reaction Time* (**2**) (Fig. 15).—The *S* keeps his finger on a button, depressing it ($\frac{1}{8}$ in.) as rapidly as possible in response to a single amber light before him. A click provides him with a ready signal before each light stimulus is presented with the foreperiod (between click and light) varying in a random order from .5 to 1.5 sec. Score is the cumulated reaction time for a series of 20 reactions.

17. *Auditory Reaction Time* (**2**) (Fig. 15).—Same procedure and scoring as Visual Reaction Time except that *S* responds to buzzer when it sounds instead of a light.

18. *Jump Visual Reaction Time* (Fig. 15).—Same procedure and scoring as Visual Reaction Time except *S* does not keep his finger on the response button, but keeps it on a cross 6 in. from the response button. He must move his hand to the button as rapidly as possible as each light stimulus appears.

19. *Jump Auditory Reaction Time* (Fig. 15).—Same as Jump Visual Reaction Time except response is to buzzer instead of light.

20. *Track Tracing* (**2**) (Fig. 16).—The *S* is required to negotiate an irregular slot pattern with a T-shaped stylus held at arms length. Score is the number of errors (contacts with the back, top and sides of slot) during four attempts. No time limits.

21. *Steadiness-Precision* (**2**) (Fig. 17).—The *S* moves a long stylus forward, slowly and steadily, away from his body, trying not to hit the sides of a cylindrical passage. Score is the number of errors (contacts with sides) during eight attempts, four through each of two passageways. No time limits.

22. *Rudder Control—Center Target* (**13**) (Fig. 18).—The *S* sits in a mock airplane cockpit, which he attempts to keep lined up steadily with one of three target lights as they come on in front of him. His own weight throws the seat off balance unless he applies and maintains proper correction by means of foot pedals. During this condition, only the center target was used. Score was the total time the cockpit is held lined up with the center light during three 30-sec. trials separated by 15-sec. rests.

23. *Rudder Control—Triple Target* (Fig. 18).—Same apparatus as 22 above, but in this condition *S* must utilize appropriate pedal control to shift the cockpit from one light to the other as these come on at random intervals. Score is the total time the cockpit is lined up with the proper light during three 112-sec. trials separated by 30-sec. rests.

24. *Printed Discrimination Reaction Time* (**2**) (Fig. 19).—Going from item to item as rapidly as possible, the *S* makes a check mark in one of four slots (arranged in an up-down, right-left pattern), according to the configuration of black and white dots. Score is the number of corrects minus twice the number wrong for one continuous 100-sec. trial.

Variables 25–31 represent different conditions administered on the Multidimensional Pursuit Apparatus (**14**) (Fig. 20). In the original version of this apparatus, *S* is confronted with four dials representing airspeed, altitude, bank, and turn indicators. Pointers on each of these dials vary irregularly, continuously and independently of one another. The *S* attempts to keep all the pointers in the center of their respective scales by compensatory adjustments of simulated stick, rudder, and throttle controls. Score is the amount of time the pointers on all four dials are lined up *simultaneously* in a zero position. In the present study, seven different tasks were performed on this apparatus. The apparatus was modified in order to allow *S* to practice on single components or on combinations of components. The following conditions were administered to *S* in the order described. Within each condition *S* received four 1-min. trials, separated by 15-sec. rests. Between conditions, *S* received a 5-min. rest, during which he was given instructions for the condition to follow.

25. *Multidimensional Pursuit—Bank.*—All dials were covered, except the lower center dial indicating "Bank." The *S* attempted to keep the pointer centered on this one dial through right and left compensatory movements of the stick control only. Score was the amount of time the pointer was "on target."

26. *Multidimensional Pursuit—Heading.*—All dials were covered, except upper center dial indicating "Heading." The *S* attempted to keep the pointer on this dial centered through compensatory right and left foot adjustments on the rudder pedals. Score was the amount of time the pointer was "on target."

27. *Multidimensional Pursuit—Air Speed.*—All dials were covered, except the extreme right dial indicating "Air Speed." The *S* attempted to keep this pointer centered by compensatory forward and backward movements of the small "throttle control" handle manipulated by his left hand. Score was the amount of time this pointer was "on target."

28. *Multidimensional Pursuit—Bank and Altitude.*—Two dials were uncovered and *S* had to keep both pointers centered through coordinated use of the stick in two dimensions. Forward and backward stick movement controlled the pointer on the altitude dial, while right and left movement controlled the pointer on the bank dial. Score was the amount of time *both* pointers were "on target" simultaneously.

TABLE 1

MEANS, *SD*'S, AND RELIABILITIES OF THE EXPERIMENTAL TASKS

Task	Mean	*SD*	Reliability[a]
1. Two-Hand Coordination	.93	.35	.80
2. Rotary Pursuit	.36	.20	.81
3. Complex Coordination	14.50	3.36	.82
4. Plane Control	.94	.38	.88
5. Pursuit Confusion—T.O.T.	1.58	.58	.82
6. Pursuit Confusion—Errors	1.13	.52	.84
7. Discrimination Reaction Time	.83	.21	.75
8. Motor Judgment	1.21	.27	.76
9. Visual Coincidence	33.0	8.05	.74
10. Two-Hand Pursuit	2.01	.51	.83
11. Single Dimension Pursuit	1.56	.29	.76
12. Rate Control	1.74	.42	.69
13. Dial Setting	22.61	3.92	.80
14. Control Adjustment	105.5	11.8	.91
15. Rotary Aiming	339.8	46.6	.92
16. Visual Reaction Time	.086	.015	.82
17. Auditory Reaction Time	.082	.012	.81
18. Jump Visual Reaction Time	.145	.018	.83
19. Jump Auditory Reaction Time	.141	.018	.84
20. Track Tracing	57.76	25.61	.88
21. Steadiness-Precision	190.7	68.76	.87
22. Rudder Control—Single Target	.59	.31	.70
23. Rudder Control—Triple Target	4.03	2.15	.82
24. Printed Discrimination Reaction Time	64.25	18.98	.87
25. MP—Bank	3.75	.21	.75
26. MP—Heading	3.67	.20	.82
27. MP—Air Speed	3.65	.25	.86
28. MP—Bank and Altitude	1.94	.45	.90
29. MP—Bank and Air Speed	2.24	.48	.96
30. MP—Bank and Heading	2.11	.47	.95
31. MP—Bank, Heading, and Air Speed	.60	.31	.89

[a] Reliabilities for Variables 1–15, 20, 21, and 25–31 are based on correlations of odd and even trials corrected for full length of the tasks. Reliabilities for Variables 16–19 and 24 are test-immediate retest correlations obtained in a previous study. Reliabilities for Variables 22 and 23 are average intertrial correlations obtained previously.

29. *Multidimensional Pursuit—Bank and Heading.*—The two dials indicating Bank and Heading were uncovered and *S* had to keep both pointers centered through simultaneous stick (right and left movement only) and foot pedal adjustments. Score was amount of time *both* pointers were "on target" simultaneously.

30. *Multidimensional Pursuit—Bank and Air Speed.*—Two dials were uncovered and *S* had to keep their respective pointers centered through simultaneous movements of the right hand (control stick right and left) and the left hand (throttle control handle to forward and back). Score was amount of time *both* pointers were "on target" simultaneously.

31. *Multidimensional Pursuit—Bank, Heading, and Air Speed.*—Three dials were uncovered and *S* had to keep their respective pointers centered through simultaneous adjustments of both hands as well as both feet. Score was amount of time *all three* pointers were "on target" simultaneously.

The reader will note the subtasks were administered in an assumed order of difficulty. The actual means obtained under the present administrative conditions are listed in Table 1. It should be noted that the "Bank" component is involved in each of the complex conditions. In turn, these involved one hand in two dimensions, one hand and feet, two hands, and finally both hands and feet.

Administrative Procedure

The complete battery of tasks was administered to 204 basic trainee airmen at Lackland Air Force Base. Half the *S*s were administered the seven subtasks of the Multidimensional Pursuit apparatus (Variables 25–31 above) before receiving the remaining tasks, while the other half received the Multidimensional Pursuit tasks last. In the administration of these other experimental tasks (Variables 1 to 24) a rotational procedure was used in which

TABLE 2

INTERCORRELATIONS[a]

Variable	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1. Two-Hand Coordination	13	24	33	13	18	25	26	22	31	14	26	21	22	04	08	05	21	23	14	21	46	43	10	06	10	07	25	09	17	13
2. Rotary Pursuit		34	23	21	22	17	28	04	21	18	19	24	32	26	09	08	12	09	27	05	29	36	14	16	23	16	21	04	19	21
3. Complex Coordination			28	11	16	32	20	05	34	21	16	32	30	06	−03	−02	02	02	26	12	28	30	17	22	23	16	34	29	35	28
4. Plane Control				16	29	25	16	13	32	15	16	16	26	02	−01	09	03	12	16	07	34	31	17	20	12	09	28	13	24	18
5. Pursuit Confusion—T.O.T.					17	25	30	13	21	08	14	26	17	01	−04	−02	06	05	09	16	17	20	18	10	07	21	24	20	15	23
6. Pursuit Confusion—Errors						22	20	−01	23	02	24	12	10	06	−00	16	09	16	20	14	29	22	14	04	−03	08	06	−01	11	08
7. Discrimination Reaction Time							19	30	27	26	20	42	26	09	09	15	22	24	10	10	26	24	53	22	27	33	29	30	36	34
8. Motor Judgment								15	23	15	29	33	32	11	03	11	08	13	17	14	34	33	15	02	10	09	22	09	16	10
9. Visual Coincidence									23	06	08	27	15	12	04	20	14	18	10	13	19	14	15	07	18	11	19	09	13	15
10. Two-Hand Pursuit										25	24	32	33	05	17	19	14	20	27	19	28	25	26	12	04	06	28	14	22	17
11. Single Dimension Pursuit											12	24	34	11	10	10	14	08	21	−05	20	15	30	22	19	15	27	26	36	20
12. Rate Control												14	19	07	13	08	14	18	01	−06	26	33	10	12	18	29	19	15	26	32
13. Dial Setting													37	17	10	09	09	08	14	15	22	20	30	08	15	19	28	19	24	25
14. Control Adjustment														14	06	06	13	04	23	15	28	29	22	09	08	03	26	14	23	14
15. Rotary Aiming															23	17	34	33	17	03	09	03	28	05	07	14	17	11	17	10
16. Visual Reaction Time																48	46	40	09	15	03	06	18	−04	05	09	06	03	10	08
17. Auditory Reaction Time																	39	47	12	21	09	12	14	02	07	10	09	−04	10	−02
18. Jump Visual Reaction Time																		71	16	15	09	11	21	06	05	05	08	07	11	04
19. Jump Auditory Reaction Time																			07	18	19	18	15	10	12	19	19	15	19	14
20. Track Tracing																				33	16	10	23	04	05	03	08	01	14	06
21. Steadiness-Precision																					09	13	14	01	07	09	12	10	16	05
22. Rudder Control—Single Target																						76	13	21	13	10	27	14	23	16
23. Rudder Control—Triple Target																							13	19	14	05	28	10	22	19
24. Printed Discrimination Reaction Time																								14	14	13	19	22	28	24
25. MP—Bank																									49	32	46	46	49	46
26. MP—Heading																										59	40	56	58	54
27. MP—Air Speed																											45	58	63	61
28. MP—Bank and Altitude																												69	64	63
29. MP—Bank and Heading																													76	70
30. MP—Bank and Air Speed																														70
31. MP—Bank, Heading, and Air Speed																														

[a] Decimals omitted.

TABLE 3

CENTROID FACTOR LOADINGS[a]

Variable	Factors										h^2
	I	II	III	IV	V	VI	VII	VIII	IX	X	
1. Two-Hand Coordination	45	−23	17	10	−13	15	05	12	−18	09	39
2. Rotary Pursuit	44	−13	23	05	12	−28	09	−29	18	−04	48
3. Complex Coordination	47	08	32	−09	14	−06	−13	−11	−06	03	40
4. Plane Control	42	−08	28	09	06	11	−26	12	09	−17	40
5. Pursuit Confusion—T.O.T.	34	−04	13	−14	−23	−11	12	13	02	−19	29
6. Pursuit Confusion—Errors	31	−24	14	13	−06	−18	−21	06	08	−22	33
7. Discrimination Reaction Time	58	04	−08	−20	−28	25	−14	−15	13	−16	61
8. Motor Judgment	42	−24	22	−08	−15	−15	18	07	03	08	38
9. Visual Coincidence	32	−09	−10	−14	−18	20	−07	05	06	09	23
10. Two-Hand Pursuit	50	−22	18	−15	11	09	−15	25	11	06	47
11. Single Dimension Pursuit	40	09	11	−14	26	17	07	−15	13	13	36
12. Rate Control	40	−04	10	28	−16	−07	15	11	20	06	36
13. Dial Setting	49	−06	08	−36	−16	06	13	−03	12	10	45
14. Controls Adjustment	45	−15	27	−25	16	03	14	−06	06	14	43
15. Rotary Aiming	30	−14	−26	−05	19	−12	20	−20	05	−12	33
16. Visual Reaction Time	27	−24	−46	10	19	04	09	03	12	17	44
17. Auditory Reaction Time	31	−30	−43	12	13	09	−11	12	17	16	49
18. Jump Visual Reaction Time	39	−33	−51	10	23	08	13	−06	−13	−13	64
19. Jump Auditory Reaction Time	46	−27	−53	25	15	09	07	09	−13	−15	71
20. Track Tracing	33	−21	05	−21	16	−27	−27	−14	−09	12	41
21. Steadiness-Precision	28	−19	−11	−15	−04	−17	−26	13	−23	14	34
22. Rudder Control—Single Target	54	−25	38	30	−14	19	05	−11	−21	06	71
23. Rudder Control—Triple Target	53	−26	37	36	−15	17	11	−12	−19	07	73
24. Printed Discrimination Reaction Time	46	−03	−14	−36	−06	10	−07	−23	12	−17	48
25. MP—Bank	41	44	06	15	12	10	−09	−10	−03	−10	44
26. MP—Heading	48	51	−11	17	−08	−08	−08	−15	03	16	60
27. MP—Air Speed	49	48	−20	12	−21	−27	04	08	12	08	67
28. MP—Bank and Altitude	64	39	08	−04	12	04	13	25	−13	−08	69
29. MP—Bank and Air Speed	55	65	−09	−08	02	−04	13	16	−19	−04	82
30. MP—Bank and Heading	69	54	−06	03	08	−09	−04	06	−05	11	81
31. MP—Bank, Heading, and Air Speed	59	54	−03	08	−09	−10	09	10	06	−08	69

[a] Decimals omitted.

the order of occurrence of each task in the series was fixed (in the order listed above), but different *S*s started at different points in the series. Two models of each apparatus were used in order to test two *S*s simultaneously. In each case, the two test models were wired into a single scoring and control console monitored by *E*.

Data Analysis Procedures

Table 1 summarizes the means, *SD*'s and reliability estimates for the task variables as administered in the present study.

For each task the obtained distributions of raw scores were transformed to normalized distributions of standard scores (stanines), each with a range from 1 to 9, a mean of 5, and *SD* of 2. Conversions were made so that the 9 end of the scale was always indicative of "good" performance (e.g., low errors, high time-on-target, low reaction time).

Since two models of each apparatus were used, an evaluation was made of possible differences between mean scores on each task variable as a function of apparatus model. *t* tests for only two variables proved significant beyond the 5% level and these were the Plane Control and Two-Hand Pursuit tasks. Consequently, for these two tasks, separate standard score conversions were made for each apparatus.

Thus, in terms of the standard score achieved, *S*'s performance was made independent of the particular apparatus used.

Results and Discussion

Pearson product-moment correlations among the 31 variables were obtained. The resulting correlation matrix is presented in Table 2. Factors were extracted from this matrix by Thurstone's Centroid Method (**15**). Extractions were continued beyond the point where the product of the two highest centroid loadings became less than the standard error of the original correlation of the same two variables. Ten centroid factors were extracted. Orthogonal rotations of the primary axes were made by means of Zimmerman's graphical method (**16**) using the criterion of simple structure and positive mani-

TABLE 4

Rotated Factor Loadings[a]

Variable	Factors[b]										h^2
	I RO	II FS	III RT	IV SAM	V AHS	VI MLC	VII RC	VIII D	IX WT	X Res	
1. Two-Hand Coordination	15	25	13	02	06	33	32	18	−12	−09	37
2. Rotary Pursuit	−09	50	−08	17	18	00	05	26	08	28	48
3. Complex Coordination	09	35	06	21	27	30	07	21	19	08	43
4. Plane Control	12	02	−01	11	25	41	19	21	03	25	40
5. Pursuit Confusion—T.O.T.	23	12	−04	01	11	−07	37	24	02	04	28
6. Pursuit Confusion—Errors	−04	04	05	06	36	11	19	31	−07	18	31
7. Discrimination Reaction Time	67	19	11	03	19	09	03	18	01	18	61
8. Motor Judgment	07	40	03	01	11	03	40	15	06	06	37
9. Visual Coincidence	36	12	14	−03	08	12	12	15	−10	01	23
10. Two-Hand Pursuit	18	13	07	18	26	32	37	00	−07	26	47
11. Single Dimension Pursuit	24	27	−02	23	05	25	−06	−09	14	26	37
12. Rate Control	−03	30	29	−16	−02	17	30	06	14	08	35
13. Dial Setting	43	40	−03	05	08	−06	27	−08	−03	13	45
14. Control Adjustment	14	46	−12	24	06	18	22	−11	−04	17	43
15. Rotary Aiming	07	26	17	38	08	−22	−02	18	04	10	34
16. Visual Reaction Time	01	15	56	24	02	−08	02	−08	−03	11	42
17. Auditory Reaction Time	06	07	63	18	10	04	06	−08	−11	19	51
18. Jump Visual Reaction Time	12	12	54	54	−01	−10	02	18	−09	−05	66
19. Jump Auditory Reaction Time	13	04	64	44	00	−02	13	27	00	−06	71
20. Track Tracing	00	29	09	24	50	00	−03	−09	−02	00	41
21. Steadiness-Precision	06	10	19	09	43	00	17	−09	−05	−19	32
22. Rudder Control—Single Target	08	44	06	01	03	52	19	44	−13	−09	73
23. Rudder Control—Triple Target	04	48	08	−06	−03	48	20	46	−13	−09	75
24. Printed Discrimination Reaction Time	52	24	01	24	21	−08	−04	06	−02	19	48
25. MP—Bank	24	07	10	12	02	21	−12	17	51	05	45
26. MP—Heading	24	27	20	−14	10	10	−10	11	61	02	60
27. MP—Air Speed	23	18	22	−19	13	−11	17	12	64	04	65
28. MP—Bank and Altitude	32	15	02	24	00	26	37	09	56	−03	71
29. MP—Bank and Heading	41	11	02	13	−02	08	22	02	75	−15	84
30. MP—Bank and Air Speed	31	24	15	09	15	19	14	09	72	04	79
31. MP—Bank, Heading, and Air Speed	33	17	10	−01	05	05	23	21	66	08	69

[a] Decimals omitted.
[b] Factors are identified as follows: I, Response Orientation; II, Fine Control Sensitivity; III, Reaction Time; IV, Speed of Arm Movement; V, Arm Hand Steadiness; VI, Multiple Limb Coordination; VII, Rate Control; VIII, Doublet; IX, Within Task Factor; X, Residual.

fold. The accuracy of the rotational procedure was checked through reproduction of original correlation coefficients and through a comparison of communalities in the centroid and rotated matrixes.

Interpretation of Factors

Factors were interpreted for psychological meaningfulness from the projections of the task variables on the rotated axes. The 10 factors identified will be described in turn. Tasks with loadings of .30 or higher are listed under each factor.

Factor I appears to be the *Response Orientation* factor previously identified (**5, 7, 11**). It has been found general to visual-discrimination reaction psychomotor tasks involving rapid directional discrimination and orientation of movement patterns.

No.	*Variable*	*Loading*
7	Discrimination Reaction Time	.67
24	Printed Discrimination Reaction Time	.52
13	Dial Setting	.43
29	Multidimensional Pursuit—Bank and Heading	.41
9	Visual Coincidence	.36
31	Multidimensional Pursuit—Bank, Heading, and Air Speed	.33
28	Multidimensional Pursuit—Bank and Altitude	.32
30	Multidimensional Pursuit—Bank and Air Speed	.31

This factor has been found distinct from the factor called Spatial Orientation, which represents the ability to comprehend the arrangement of a visual stimulus pattern. The present factor appears to involve the ability to make the correct movement in relation to the correct stimulus. In other words, "Given this stimulus, which way should I move?" The Discrimination Reaction Time tests were among the best measures of this factor in previous studies (**5, 7, 11**). In the Dial Setting test, critical features involve moving the correct one of four knobs in correct relation to the four stimulus windows, but even more crucially turning the knobs in the correct direction so as to achieve a match more quickly.

The lower loading (.23) of the Pursuit Confusion test on this factor may at first appear disturbing. This test involves mirror tracing, and the "which way do I move" question would certainly appear critical. Similarly, the Two-Hand Coordination and Two-Hand Pursuit tests involve such response decisions. The distinction between these three tests and those on the factor appears to be that these involve continuous responses to continuous stimuli. Tests on the factor involve discrete responses to successive stimuli.

The presence of the Visual Coincidence test on this factor implies that this factor may extend to situations in which the choice is whether to respond or not. Of special interest is the presence of four subtasks of the Multidimensional Pursuit task on this factor. It is to be noted that these are the subtasks involving more than one component at a time. Here associating the correct control with the correct dial is especially critical in obtaining a score. The three subtasks involving no choice of control did not load on this factor.

Factor II is defined as *Fine Control Sensitivity*.

No.	*Variable*	*Loading*
2	Rotary Pursuit	.50
23	Rudder Control—Triple Target	.48
14	Control Adjustment	.46
23	Rudder Control—Single Target	.44
13	Dial Setting	.40
8	Motor Judgment	.40
3	Complex Coordination	.35
12	Rate Control	.30

All of these tasks involve the ability to make fine, highly controlled (but not overcontrolled) adjustments at some critical stage of performance. These results confirm previous indications (**11**), that this ability extends to arm-hand as well as to leg movements. The

presence of Dial Setting and Rate Control suggests the extension to wrist-finger movements as well. Of special interest is the substantial loading of the Control Adjustment task on this factor. This device was specifically constructed as a measure of this hypothesized factor. On the negative side, Pursuit Confusion, originally hypothesized as a measure of this factor, loaded on another factor.

Factor III is restricted to the four reference measures of *Reaction Time.*

No.	*Variable*	*Loading*
19	Jump Auditory Reaction Time	.64
17	Auditory Reaction Time	.63
16	Visual Reaction Time	.56
18	Jump Visual Reaction Time	.54

None of the more complex tasks appear on this factor. This factor has been defined simply as the speed with which *S* can react to a stimulus when it appears. These results agree with previous indications (**2, 3, 10**) that individual differences in this ability are independent of whether the stimulus is auditory or visual. These results also confirm that this Reaction Time factor does not contribute to individual differences in the Discrimination Reaction Time task briefly practiced. Reaction Time does contribute variance, however, after considerable practice is given on this task (see **10**).

Factor IV is restricted to those tasks in the present analysis which have previously identified a *Speed of Arm Movement* factor. This represents simply the speed with which *S* can make a gross, discrete arm movement.

No.	*Variable*	*Loading*
18	Jump Visual Reaction Time	.54
19	Jump Auditory Reaction Time	.44
15	Rotary Aiming	.38

The distinction between this factor and the Reaction Time factor is again brought out. For example, this Speed of Arm Movement contributes variance in those reaction-time tasks which require a 6-in. arm movement to the response button (Variables 18, 19) but not in those reaction time tests requiring only a button-pressing response. The Rotary Aiming task involves no sudden onset of stimuli but does involve a series of rapid, discrete arm movements. This factor is not found in any of the complex tasks as represented in the present study. Previous work has shown that this factor does contribute variance in performance on the Complex Coordination, Discrimination Reaction Time, Plane Control, and Rotary Pursuit tasks, but only at high levels of proficiency achieved after continued practice on these tasks (**3, 5, 9, 10**).

Factor V is identified as *Arm-Hand Steadiness*, and defined as the ability to make precise and steady arm-hand movements of the type which minimize strength and speed.

No.	*Variable*	*Loading*
20	Track Tracing	.50
21	Steadiness-Precision	.43
6	Pursuit Confusion (Errors)	.36

This factor has appeared in a number of our studies. It has been found to extend to tasks requiring maintenance of a steady arm position, to tasks requiring steadiness during an arm movement, and to various planes of movement, etc. (**2, 6, 12**). However, in no previous study has it been found to extend to more complex psychomotor performances. Of interest in the present study is the loading of Pursuit Confusion (Errors) on this factor. It will be recalled that this score represents hits against the sides of a depressed slot.

Factor VI includes tasks, all of which involve simultaneous manipulation of multiple limbs.

No.	*Variable*	*Loading*
22	Rudder Control—Single Target	.52
23	Rudder Control—Triple Target	.48
4	Plane Control	.41
1	Two-Hand Coordination	.33
10	Two-Hand Pursuit	.32
3	Complex Coordination	.30

This confirms previous indications (**11**) that the ability represented by this

factor is general to tasks requiring coordination of two feet (e.g., Rudder Control), two hands (Two-Hand Pursuit, Two-Hand Coordination), and hands and feet (Plane Control and Complex Coordination). This factor is labeled *Multiple Limb Coordination.* These are the only tasks in the study which require multiple limb involvement, except for certain subtasks on the Multidimensional Pursuit Apparatus. It should be noted that all the tasks loaded on this factor present a display which allows *S* to assess the degree to which he is "coordinating." In the Multidimensional Pursuit task there is no single indication of this, but instead *S* must divide his attention between spatially separate dials. It is possible that these differences in control-display characteristics are of consequence with respect to the abilities measured by such tasks.

Factor VII appears the same as that called *Rate Control* in one of our recent studies (**11**). Since loadings are low, this must be regarded as our most tentative interpretation. However, all these loadings are consistent with our interpretation.

No.	*Variable*	*Loading*
8	Motor Judgment	.40
5	Pursuit Confusion (Time on Target)	.37
10	Two-Hand Pursuit	.37
28	Multidimensional Pursuit—Bank and Altitude	.37
1	Two-Hand Coordination	.32
12	Rate Control	.30

The common feature of all these tasks is that there is an element of pursuit involved. Each task requires the examiner to make anticipatory adjustments relative to changes in speed and/or direction of a continuously moving object. All compensatory and following pursuit tasks in the present study, with the exception of the Single Dimension Pursuitmeter, are located on this factor. These results confirm our previous indications that this factor cuts across the traditional categorization of pursuit tasks into "following pursuit" (e.g., Rate Control and Two-Hand Coordination) and compensatory pursuit (e.g., Two-Hand Pursuit). These results suggest that from the point of view of individual differences, this distinction may be arbitrary, and the nature of such tasks may better be thought of in terms of a third underlying variable. Our previous study showed this factor to be general to tasks other than those traditionally called pursuit. The presence of the Motor Judgment task on this factor in the present analysis is further evidence of this. This task was included specifically to test the original definition of this factor. This is not a typical pursuit task, but there is considerable involvement of rate judgment and adjustment. The *S* must consider the rate of movement of two different disks, a pointer, and of his own control movement to make an adequate response. It will be noted that the task, Visual Coincidence, does not load on this factor. This task was included to see if judgment of stimulus rate as opposed to response rate was the key to this factor. It will be recalled that only a button-pressing response is required in this task. The absence of this task from Factor VII suggests that this factor represents skill in the motor aspects of the response rather than skill in the purely interpretational aspects of the rate of the stimulus.[2]

Factor VIII is a doublet factor confined to the subtasks on the Rudder Control Task and is of doubtful psychological significance.

Factor IX is confined to the seven subtasks administered on the Multidimensional Pursuit device and does not extend to the other ability measures. Moreover, there is no consistent patterning of these loadings from one such task to another (see Table 4) or between the three clear-cut levels of difficulty among these seven subtasks evidenced in Table 1. It is possible that this "within-task" factor represents specific or spurious

[2] This has been confirmed in a recent, still unpublished, study in which we obtained only insignificant correlations between psychomotor pursuit tasks and motion picture tests of rate judgment and discrimination.

variance due to experimental dependence of an unknown source. Although scores on each condition were obtained independently at successive stages, they were obtained on the same apparatus in the same order and during a single testing period for each *S*. It is not known, for example, to what extent positive or negative transfer effects might be represented in the intercorrelations among these subtasks. Whatever the nature of Factor IX, it is clear that we cannot define it in terms of the more general ability measures included in the present study.

It was hoped that inclusion of the several scores on the Multidimensional Pursuit Apparatus would yield information on the relation between task complexity (in terms of number of display-control components involved) and factorial complexity (number of abilities required). As indicated earlier, the main conclusion possible from these data is that Response Orientation is involved when the task is complicated by the requirement that two or three controls be operated, but that this factor is not involved in the single component subtasks. Perhaps just as important is the finding that these results are independent of the particular control used or combination of controls (two feet, two hands, hand and feet, etc.) involved. The critical skill appears to be that of associating the correct control movement with the various display variations.

It should also be pointed out that a large portion of the variance in the other experimental tasks remains specific to individual tasks. However, this should not minimize the importance of discovering whatever common variance there is. Generalization of experimental results from one task to another might well depend on this common variance. However, this remains to be evaluated in other laboratory contexts.

The importance of discovering this common variance certainly applies to the kind of predictions one may be able to make about individual differences in psychomotor performances. For a recent review of the utility of these concepts to the description and prediction of individual differences of practical consequence, the reader is referred elsewhere (**4**).

It is not possible here to review the place and limitations of factor analysis in experimental psychology. It would seem, however, that results of such studies can provide a set of working dimensions for conceptualizing task variables in terms of common ability requirements. The utility of the descriptive models provided by this approach needs to be assessed by nonfactorial investigations.

Summary

The study confirms previous indications that movement reaction tasks of the kinds investigated may be grouped into several broad classes representing common ability requirements. The ability categories inferred from the factor analysis results were labeled Fine Control Sensitivity, Multiple Limb Coordination, and Response Orientation. A factor called Rate Control was considered more tentative. Definitions of these factors and the diverse kinds of tasks to which they apply are described. Three other factors were identified as Arm-Hand Steadiness, Reaction Time, and Speed of Arm Movement, but these did not contribute to performance in the more complex movement reaction tasks under the present conditions of administration.

References

1. Brown, J. S., & Jenkins, W. O. An analysis of human motor abilities related to the design of equipment and a suggested program of research. In P. M. Fitts (Ed.), *Psychological research on equipment design*. Washington D. C.: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Res. Rep. No. 19.)
2. Fleishman, E. A. Dimensional analysis of psychomotor abilities. *J. exp. Psychol.*, 1954, **48**, 437–454.
3. Fleishman, E. A. Predicting advanced levels of proficiency in psychomotor skills. In G. Finch & F. Cameron (Eds.), *Symposium on Air Force human engineering, personnel, and training research*. Washington: National Academy of Sciences—National Research Council, Publ. 455, 1956. Pp. 142–151.

4. Fleishman, E. A. Psychomotor selection tests: research and application in the U. S. Air Force. *Personnel Psychol.* 1956, **9,** 449–467.
5. Fleishman, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, **41,** 263–272.
6. Fleishman, E. A. An analysis of positioning movements and static reactions. *J. exp. Psychol.*, 1958, **55,** 13–24.
7. Fleishman, E. A. Factor structure in relation to task difficulty in psychomotor performance. *Educ. psychol. Measmt*, 1957, **17,** 522–532.
8. Fleishman, E. A., & Hempel, W. E., Jr. A factor analysis of dexterity tests. *Personnel Psychol.*, 1954, **7,** 15–32.
9. Fleishman, E. A., & Hempel, W. E., Jr. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19,** 239–252.
10. Fleishman, E. A., & Hempel, W. E., Jr. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49,** 301–312.
11. Fleishman, E. A., & Hempel, W. E., Jr. Factorial analysis of complex psychomotor performance and related skills. *J. appl. Psychol.*, 1956, **40,** 96–104.
12. Hempel, W. E., Jr., & Fleishman, E. A. A factor analysis of physical proficiency and manipulative skill. *J. appl. Psychol.*, 1955, **39,** 12–16.
13. Melton, A. W. (Ed.) *Apparatus tests.* Washington D. C.: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Program Res. Rep. No. 4.)
14. Melton, A. W. (Ed.) *Apparatus tests (supplement).* Washington D. C.: U. S. Government Printing Office, 1947. (AAF Aviat. Psychol. Program Res. Rep. No. 4 Supplement.)
15. Thurstone, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1947.
16. Zimmerman, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11,** 51–55.

(Received May 25, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

THE DIMENSIONAL ANALYSIS OF A NEW SERIES OF FACIAL EXPRESSIONS

TRYGG ENGEN, NISSIM LEVY, AND HAROLD SCHLOSBERG

Brown University

In a series of three papers (**6, 7, 8**) Schlosberg showed that photographs of facial expressions of emotions could be described very well in terms of three dimensions: (*a*) Pleasantness-Unpleasantness (P-U), (*b*) Attention-Rejection (A-R), and (*c*) Sleep-Tension (S-T). The resulting figure is an irregular cone, roughly comparable to the bottom half of the color solid, with P-U and A-R corresponding to the Blue-Yellow and Red-Green axes, while S-T, the intensive dimension, corresponds to the line from Black to Mid-Gray. These studies were made on two sets of photographs, the extensive male series of Frois-Wittmann (**3**) and the shorter and less adequate female series of Ruckmick (**5**). The basic method of locating the photos on the figure was to have them rated on a 9-point scale for each of the three dimensions.

Engen and Levy (**1**) obtained results providing surprisingly good confirmation of the dimensional values of selected pictures of the Frois-Wittmann series. They used the constant-sum method of judgment. This method is sometimes claimed to give an equal-interval scale with a true zero; at the very least the method has the advantage of providing checks on the internal consistency of the scale values obtained.

Neither of the series used in the previous experiments was adequate for a complete analysis of the Sleep-Tension dimension, for the pictures were intended to represent *Emotions*, a traditional category which stresses only the tense or highly activated states of the organism. Therefore, when an opportunity was presented to make a new series of pictures of a female face, the present authors were careful to include some relaxed expressions. This new series has the further advantage of modern technical advances such as electronic flash photography which enabled us to obtain sharp pictures even during movement (**2**). The photographs comprising this new series, together with some norms, have been reproduced in Morgan's laboratory manual (**4**). Some results from a lay group and from undergraduates are presented elsewhere (**2**). The availability of this improved series of pictures stimulated the present writers to study the effect of certain variables on the scale values of the pictures.

General Method

All of the data reported in this paper were obtained from students in laboratory periods of one of the courses in elementary psychology at Brown University. The sections ranged from 25 to 40 students in size and contained both sexes with a ratio of two men to each woman. Since the experiments were done toward the middle of the term, *Ss* were familiar with the general routine of individual and group experiments. The pictures to be rated were always presented singly, on 2 × 2-in. slides, under moderately dim illumination. Each *S* sat in a tablet armchair, placed to give adequate view of the projected pictures. Judgments were recorded by *Ss* on specially prepared forms.

Before the start of each experiment, the general method was explained and illustrated by having *Ss* assign ratings of from 1 to 9 to each of two pictures representing the extremes of the dimension to be judged first. These "an-

chors" were from the Ruckmick series[1] and were presented as a pair to bring out the differences between them. The pictures to be rated were then presented seriatim at a rate that permitted *Ss* to make and record careful judgments but not to develop indecision; after the first few pictures, the rate was about six per minute.

After the selected series of pictures had been rated on one dimension, *Ss* folded their record sheets to cover the ratings. Another dimension was then discussed, again with the aid of anchoring photos, and a series of pictures was presented for rating on the new dimension. This general procedure was continued for about an hour, leaving adequate time for *Ss* to work up some group results before the close of the 2-hr. period.

The fact that there were a number of separate laboratory groups permitted variation of such experimental factors as the order in which the dimensions were judged and the length and constitution of the series of pictures exposed. The experiments were done over a period of two years. No student participated in more than one experimental session.

Results and Discussion

Experiment I

Ninety-six *Ss* judged all 48 photographs of the new series on all three dimensions. The medians for each picture on each dimension are presented in Table 1. The interquartile ranges are presented because the distributions of judgments in the end categories were too skewed to justify the calculation of *SD*'s. It will be noted that the pictures spread very well to the ends of each dimension; i.e., *Ss* used the whole scale for each dimension. The variability of ratings, as measured by the interquartile range, is in general smallest for the P-U dimensions, largest for the A-R dimensions, with S-T being slightly less variable than A-R.

Experiment II

Six groups of 35–40 *Ss* participated in this study. Sixteen pictures were chosen and *Ss* were asked to rate these pictures on each of the three dimensions. The order of presentation of the dimensions was varied from group to group so that each dimension occupied each ordinal position in two different groups. Thus it was possible to determine whether or not practice on judging one dimension would alter the ratings assigned to another dimension. Ten of the 16 pictures were selected at random for each analysis, and the average scale value for each picture was determined for each of the three dimensions for each group of *Ss*. There was a slight suggestion that practice tended to make the ratings of the pictures move slightly toward the midpoint of the dimension when it was presented after one or two other dimensions. However, the shifts were small and somewhat inconsistent; none was greater than .5 scale point for P-U, 2 of the 10 were as great as .6 point for A-R, and three were about a full-scale step in S-T. It should be noted that two of the three largest shifts appeared in pictures near the middle of the S-T dimension, and they were not consistent with the trend toward conservative ratings mentioned above. Hence, it seems reasonable to assume that all shifts in scale values are random variations and not the result of previous judging experience.

After *Ss* of these six groups had judged all 16 pictures on each of the three dimensions, they had a rest of about 45 min. and then repeated the process, maintaining the same order of presentation. Fresh data sheets were used to encourage judgments independent of the original ones. The two sets of ratings made it possible to get some measure of the reliability of individual *Ss*. Correlations were calculated between first and second ratings on each dimension

[1] Ruckmick Picture 32 was used to represent Pleasantness; No. 8, Unpleasantness; No. 7, Attention; No. 13, Rejection; No. 26, Sleep; and No. 4, Tension.

TABLE 1

MEDIANS AND INTERQUARTILE RANGES OF THE 48 PICTURES OF THE LIGHTFOOT SERIES OF FACIAL EXPRESSIONS ON THREE DIMENSIONS: PLEASANTNESS-UNPLEASANTNESS (P-U), ATTENTION-REJECTION (A-R), AND SLEEP-TENSION (S-T)

(N = 96)

Picture Number	P-U		A-R		S-T	
	Median	Q_3-Q_1	Median	Q_3-Q_1	Median	Q_3-Q_1
1	8.1	1.27	8.1	1.56	7.2	3.77
2	3.8	1.36	6.8	3.04	4.5	2.91
3	4.9	1.59	4.6	2.22	4.7	1.82
4	6.4	1.27	6.6	1.57	5.8	1.59
5	3.0	1.54	4.6	2.99	3.7	2.45
6	1.3	1.00	8.6	1.53	8.6	1.10
7	3.8	2.31	4.2	2.17	4.1	3.40
8	6.6	1.70	6.7	1.99	6.6	1.90
9	3.9	1.66	3.0	2.02	3.3	1.75
10	1.3	1.00	8.8	1.19	8.9	.63
11	2.2	1.30	7.1	4.07	7.7	1.60
12	4.4	1.30	3.8	2.20	3.1	1.80
13	5.9	1.15	5.4	2.02	4.8	1.38
14	6.9	1.12	6.0	1.88	5.0	1.43
15	8.8	.81	7.8	1.51	7.1	2.01
16	7.0	1.59	5.9	3.09	4.0	2.65
17	5.7	1.31	4.7	1.95	4.5	1.58
18	8.5	1.12	7.9	1.03	7.4	1.75
19	8.7	2.36	8.9	.64	8.9	.65
20	3.3	1.71	2.5	2.16	3.1	1.06
21	3.7	1.27	3.8	2.30	3.4	1.70
22	2.2	1.37	5.8	5.13	7.1	1.82
24	4.3	1.33	4.6	2.69	4.6	1.71
26	6.1	.78	5.8	1.21	5.4	1.52
28	3.5	1.31	6.1	3.11	8.2	1.69
29	2.1	1.16	8.0	2.77	8.2	1.14
30	6.7	1.85	4.2	3.48	6.6	2.34
32	7.4	1.12	6.8	1.61	5.9	1.60
33	4.5	1.21	3.4	1.65	3.2	1.51
35	8.2	1.12	7.6	1.65	6.9	3.10
36	2.9	1.65	3.0	2.31	5.1	2.83
37	2.2	1.60	2.2	2.43	6.4	5.03
38	5.9	1.10	5.7	1.75	4.6	1.68
39	8.8	.67	8.0	1.49	7.8	1.90
41	3.6	1.36	3.3	1.75	4.2	2.48
42	6.1	1.08	6.2	2.13	5.1	1.99
43	6.3	1.27	6.1	1.33	5.0	1.87
45	4.1	1.35	4.2	2.69	3.4	2.76
46	4.9	1.21	4.8	1.53	4.2	1.96
47	2.6	1.43	6.3	3.73	7.2	1.46
48	6.9	.95	6.8	1.25	5.8	1.41
49	4.0	1.36	3.0	1.88	2.2	1.25
50	6.3	1.07	6.1	1.34	4.9	1.53
51	1.1	.57	8.6	4.04	8.9	.58
52	2.3	1.22	2.4	2.10	4.2	4.27
54	3.8	1.14	2.9	1.77	3.6	2.32
55	8.1	.94	8.1	1.21	7.7	1.77
56	4.1	2.47	1.3	2.42	1.0	.53

for each of the 225 *Ss* on all 16 pictures. The medians and quartiles of these correlations are listed in Table 2. It will be seen that *Ss* give surprisingly reliable ratings, especially on the P-U and S-T dimensions. The A-R dimension clearly yields the lowest reliabilities, but still gives

a median correlation in the mid-eighties.

The consistency with which *S*s rated the several pictures on each of the three dimensions is surprising in view of their reported difficulty in the task. In previous studies we have dealt only with means or medians of groups of *S*s and have attributed the stability of these measures to the cancellation of random variability among *S*s. The present results show that *S*s can make self-consistent ratings of pictures of facial expressions with a 45-min. test-retest interval. It would be of interest to determine the level of reliability after longer intervals.

The variability and reliability of ratings in the present study confirm the findings of Engen and Levy (**1**) that P-U is the most stable dimension and that A-R is least stable. However, the present study shows that S-T is almost as stable as is P-U, whereas Engen and Levy found it comparable to A-R. The most plausible explanation for this divergence in results is that the range of stimuli along the S-T dimension was considerably restricted in the Frois-Wittmann series used by Engen and Levy, whereas a deliberate effort was made to include pictures near the sleep end of the dimension in the present series. From what is known of the judgment process, restricting the range of stimuli should increase variability of ratings (**9**).

Experiment III

This experiment was designed to test the hypothesis that a series restricted to one end of the range would increase the variability of judgments despite the anchors. The test was made by comparing judgments made on the same four pictures when they were presented in a series of eight restricted to the upper half of the S-T dimension with those obtained on the same four pictures presented in a series of eight that covered the whole range of the dimension. The experimental sequence started with a practice series judged on P-U. Then followed the restricted-range S-T series. Next came a buffer series judged on A-R, followed by the full-range S-T series.

TABLE 2

MEDIANS AND QUARTILES OF PEARSON *r*'s BASED UPON 225 INDIVIDUAL *r*'s

Dimension	Q_1	Median	Q_3
P-U	.88	.94	.96
A-R	.76	.87	.93
S-T	.87	.92	.95

TABLE 3

MEDIANS AND QUARTILES OF DISTRIBUTIONS OF JUDGMENTS OF FOUR PICTURES WHEN PRESENTED IN A RESTRICTED AND A FULL-SCALE SERIES

Picture Number	Restricted			Full		
	Q_1	Median	Q_3	Q_1	Median	Q_3
4	3.56	4.83	5.84	4.08	5.07	6.05
28	5.58	6.15	6.84	6.47	7.38	7.98
51	8.08	8.78	9.15	8.48	8.84	9.16
55	4.88	6.29	7.55	5.61	7.13	8.01

Medians and quartiles of the ratings of the pictures common to the full and restricted series are presented in Table 3. The values confirm the prediction that variability is increased by restricting the range of the series to be judged. It may also be seen that the median scale values were shifted. As would be expected, Picture 51 which is at the top of both full and restricted scales, showed no appreciable shift, but Pictures 28 and 55, which are normally more than a scale division from the top of the full-range scale, dropped an additional step when presented in the restricted series. The remaining picture, No. 4, showed very little change, although it is normally about the middle of the

full scale, and might have been expected to drop considerably when shown in the restricted scale. However, this picture had little reason to show a "series effect," for it happened to be the first picture shown in both sets of pictures. This was an unfortunate oversight in planning the experiment. But the remaining data, both on variability and on absolute scale position, indicate that the scaling of the dimensions of Facial Expression, as of other psychophysical processes, is dependent on the range of objects that are presented (**9**).

Summary

In three experiments, pictures from a new series of facial expressions were presented by projection to several groups totaling about 400 undergraduate students of psychology. The *Ss*' basic task was to rate the pictures on a 9-point scale for each of Schlosberg's three dimensions: Pleasantness-Unpleasantness, Attention-Rejection, and Sleep-Tension. Experiment I, which yielded medians and quartiles for 48 pictures, showed that P-U was judged most consistently, followed by S-T and A-R.

Experiment II, involving repeated judgment on shorter series, showed no clear-cut practice or transfer effect from one dimension to another. It was also found that the reliability of judgments by single *Ss* was surprisingly high for P-U, with S-T a close second, and even the poorest dimension, A-R, had a median reliability around .85. The relatively good showing of S-T in these experiments contrasted to earlier results, presumably because the new series of pictures actually spreads from Sleep to Tension, whereas the series used in earlier experiments were restricted to relatively alert expressions. This hypothesis was strengthened by Exp. III, in which it was shown that restricting the range of pictures exposed in a series increases the variability of judgments.

References

1. Engen, T., & Levy, N. Constant-sum judgments of facial expressions. *J. exp. Psychol.*, 1956, **51,** 396–398.
2. Engen, T., Levy, N., & Schlosberg, H. A new series of facial expressions. *Amer. Psychologist*, 1957, **12,** 264–266.
3. Hulin, W. S., & Katz, D. The Frois-Wittmann pictures of facial expression. *J. exp. Psychol.*, 1935, **18,** 482–498.
4. Morgan, C. T. *Student's workbook to accompany introduction to psychology.* New York: McGraw-Hill, 1956.
5. Ruckmick, C. A. A preliminary study of the emotions. *Psychol. Monogr.*, 1921, **30,** No. 3 (Whole No. 136). Pp. 30–35.
6. Schlosberg, H. A scale for judgment of facial expressions. *J. exp. Psychol.*, 1941, **29,** 497–510.
7. Schlosberg, H. A description of facial expressions in terms of two dimensions. *J. exp. Psychol.*, 1952, **44,** 229–237.
8. Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61,** 81–88.
9. Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received May 28, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

TASK-INFLUENCE AND THE STABILITY OF GENERALIZED EXPECTANCIES[1]

JOSEPH F. RYCHLAK

Florida State University

The expectancy construct owes much of its origin to the early level of aspiration (LOA) research. Frank (**5**) demonstrated that *S*'s LOA at any given time was determined by his previous experience of success or failure on the task at hand: the greater the past success, the higher the present LOA and vice versa. Jucknat (**5**) found that when the success and failure series were made to look more different, the effects noted by Frank were less marked, although in the same direction. More recently, several investigators, operating within the framework of Rotter's Social Learning Theory (SLT) (**3; 4; 8,** pp. 105–183), have noted findings which reaffirm and enlarge upon the LOA research. Rotter has developed formulae to demonstrate that any given expectancy (E) actually has two components: generalized expectancy (GE) from related tasks, and expectancy specific to the task at hand (E′). The SLT research suggests that with increasing experience on a given task, GE loses its influence to E′ in the determination of a given E.

The present investigation deals exclusively with GE. This was accomplished by not allowing *S*s to see the tasks they were about to perform, so that they could not evaluate the specific aspects of a given task (E′), but had to rely entirely upon past experience in related tasks (i.e., GE). The experiment was designed to test two central hypotheses: (*a*) Stability of GE from one task to another is an increasing function of the frequency of reinforcement on the first task; and (*b*) when frequency and value of positive or negative reinforcement is held constant, the number of past tasks attempted by the individual serves to stabilize GE as an increasing function.

METHOD

Equipment.—The tasks were six performance tests of manual dexterity. They were numbered in the order of their experimental appearance, and each test was adapted to a 1-min. trial procedure. The better known tests and their numbers were: Minnesota Rate of Manipulation, 1; Pennsylvania Bi-Manual Worksample, 3; Purdue Pegboard, 4; and O'Conner Tweezer Dexterity, 5 (**1**). The remaining two tests were of the pegboard variety, taken from a battery used by The Adjutant General's Office, Department of the Army, to evaluate the mechanical aptitude of service personnel. Test 2 involved putting several brass washers on capped pegs, before inserting the latter in the board. Test 6 was a large, standard pegboard, with 2-in. doweled wooden pegs for insertion.

Subjects.—The *S*s were drawn from a large senior high school. All were white, American-born males. In the main, they came from manual labor economic backgrounds and only a small percentage planned to attend college. Criteria for participation in the experiment were attendance in a study hall and willingness to serve as *S*. There were few refusals. During the course of the research approximately 21 different study periods were utilized. Subjects from sophomore, junior, and senior classes were assigned randomly to experimental groups, so that each group contained equal members from each class level. In all, 120 students were used. Approximately 20% of each grade level

[1] This paper is the major portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, of the Ohio State University. The writer wishes to thank Julian B. Rotter, his advisor, and L. O. Andrews, who helped in arranging *S*s.

in school attendance during the run of the study was included.

Procedure.—The plan of the experiment was to have groups receive varying amounts of task experience preliminary to the critical experimental trial, at which time they would be given a reinforcement significantly out of proportion with their previous reinforcements in order to stimulate a change in GE. All *S*s took Tests 4, 5, and 6; however, only two of the experimental groups took Tests 1, 2, and 3 before going on to the later tests in the series. In addition to controlling the number of tasks *S* might have experience on, *E* also controlled the frequency and value of scores "earned" on the tests.

The specific instructions were uniform for each experimental condition. Subjects were told that the series of tests they were about to take had been administered to thousands of high school students all over the country. They were told that each test was a little different from the other, but that all tests measured "about the same thing, the ability to use your hands." Flicking through a booklet of tables, *E* further explained that all scores would be based upon how well other high school students of a comparable age and grade had done. Later, the pseudo-norms were obviously referred to in "deriving" each *S*'s score.

Previous research (**2; 8,** p. 177) indicated that groups of *S*s having no knowledge of a task on which the possible scores were reported to range from 0 to 50 points, would state a mean expectancy of around 28 on their first attempt. This finding was used to establish an anchor point defining a positive or negative reinforcement. That is, at the outset of the experiment, prior to actually seeing the first task, *S*s were instructed that their scores could fall between 0 and 50, and that the average person their age and grade level scored 28 points on *any single test.* Each *S* was then requested to write a covert statement of expectancy for the first test on a slip of paper and to drop it into a box placed within reach. Although introduced to *S* as of no real importance, the covert estimates were recovered following the experimental session since they represented *S*'s initial GE for tasks of this nature.

Experience in either one or four different tasks was then provided. The various pretraining conditions, with both a positive and a negative reinforcement sequence under each condition, were as follows: one trial on Test 4; four trials on Test 4; one trial on each of Tests 1, 2, 3, and 4. Six groups of 20 *S*s each were required to meet all experimental conditions. Thus, two major comparisons could be made on the basis of this pretraining: one vs. four trials in a single task, and four trials in a single task vs. four trials in four different tasks.

Groups positively reinforced in the pretraining were given 10 points above the anchor point of 28, and groups negatively reinforced 10 points below 28. For example, *S*s receiving but one reinforcement on Test 4 were given a score of either 18 (negatively reinforced) or 38 (positively reinforced). Subjects receiving four reinforcements, on the same test or on each of four different tests, all received either a series of 16, 20, 17, 19 (negatively reinforced) or 36, 40, 37, 39 points (positively reinforced).

Following pretraining, two statements of overt (i.e., *S*s stated aloud to *E*) GE were elicited. Immediately preceding Test 5 all *S*s stated an expectancy. Their score on Test 5 (controlled by *E*) was then 11 points above (for the *S*s negatively reinforced in pretraining) or 11 points below (for the positively reinforced *S*s) this Test 5 estimate. A second statement of overt expectancy was then requested for Test 6, and the difference between the two overt estimates was taken as the measure of *change* in GE.[2] On the last trial of the experimental session each *S* was reinforced at least 7 points above his second overt estimate, so that everyone ended the experiment with a success experience. Before leaving, all *S*s were asked to keep the experimental procedure confidential until *E* had left the school.

It is important to emphasize that *S* knew nothing about the specific task he was about to take at the time of making his expectancy statements. The tests were kept in cabinets, out of sight until a given administration was called for, and *S*s could not see their "next" test until actually in the process of taking it. It is in this sense that all three estimates can be viewed as statements of GE. The *S* had absolutely nothing to go on except past experience in related tasks. He was essentially being asked to state his GE for success in tasks of this nature.

Results and Discussion

Change in GE was ascertained through the absolute difference be-

[2] The 3 *S*s who made unusual shifts (i.e., raised their Test 6 expectancies after being negatively reinforced on Test 5, or lowered their Test 6 GE estimates after positive reinforcement on Test 5) were dropped from the sample. Since unusual shifts in expectancy have atypical personality correlates (**7**) beyond the scope of this investigation, it was decided at the outset to remove such *S*s from the study as they appeared.

TABLE 1

GE Statements, D Scores, and Tests of Significance
(Mean Values)

Group	Covert GE	Test 5 GE	Test 6 GE	D score	SD_D	t	
1P-S	24.85	32.10	25.75	−6.35	1.71	(1P-S & 4P-S)	1.123
1N-S	26.25	20.02	26.82	6.80	3.76	(1N-S & 4N-S*)	.303
4P-S	25.50	32.50	25.30	−7.20	2.82	(4P-S & 4P-D)	4.785**
4N-S	27.15	20.05	26.55	6.50	2.11	(4N-S & 4N-D)	5.400**
4P-D	26.05	34.50	31.25	−3.25	2.23	*	
4N-D	24.05	18.35	21.20	2.85	2.06	*	

* Group 1N-S was removed from the final analysis of variance.
** $P = .001$.

tween overt expectancy statements for Tests 5 and 6. Such difference scores (D scores) were determined for each *S* and combined into a listing according to experimental groups. Table 1, to be read from left to right, contains the mean GE statements, mean D scores, and tests of significance for all experimental conditions. The experimental groups are identified according to the number and value of pretraining reinforcements given them. For example, 1P-S means one positive reinforcement on the same test, 4N-S means four negative reinforcements on the same test, and 4P-D means four positive reinforcements on different tests.

Table 1 reveals generally smaller group means for the covert GE statements than the 28 points found in earlier studies (**2**; **8,** p. 177). The over-all *SD* of covert GE statements was 5.32. An analysis of variance on the covert GE statements failed to reach a significant *F* ratio (**6**), indicating that *S*s were successfully randomized. The original analysis of variance of the D scores yielded a value of 9.86 ($P = .001$). Note, however, that Group 1N-S has a D score variance much in excess of any other group. This fact left the possibility that an artifact influenced the *F* ratio. Consequently, to be consistent, a second analysis was done omitting Group 1N-S (which failed to differ significantly from 4N-S in any case). The second *F* ratio was more significant, with a value of 12.69. Group comparisons and *t* ratios (**6**) having relevance to the purposes of the study are listed at the extreme right of Table 1.

Interpretation of the statistical analyses supports the hypothesis that when frequency and value of positive or negative reinforcement are held constant, the number of past tasks attempted by *S* serves to stabilize GE as an increasing function. However, the proposal that stability of GE from one task to another is an increasing function of the frequency of reinforcement on the first task was not substantiated.

The latter finding may be due in large measure to the 28-point norm given *S*s. In order to differ significantly in stability of GE from an *S* receiving four pretraining reinforcements on Test 4, an *S* experiencing only one pretraining reinforcement on Test 4 would have to make a rather extreme shift in GE estimate for Test 6. However, since the latter *S* knew that the "average person" scored only 28 on any single test, it was not likely that he would move much above or below this anchor point in his GE estimate for Test 6. There is some evidence, from comparable groups, that extreme shifts in expectancy do occur when *S* is not told anything

about the usual or average score of a task (**8**, p. 177).

Summary

This study contrasts the stability of generalized expectancies (GE's) developed under two pretraining conditions: several reinforcements in one motor-skill task, with one reinforcement in each of several tasks (the frequency and value of all reinforcements being held constant). It also examines the stability of GE's based upon differing frequencies of past reinforcement in the same task. Subjects were 120 white males chosen from a senior high school.

The findings support the hypothesis that when frequency and value of positive or negative reinforcement are held constant, the number of past tasks attempted by *S* serves to stabilize GE as an increasing function. However, the hypothesis that stability of GE from one task to another is an increasing function of the frequency of reinforcement on the first task was not substantiated.

The results can be theoretically interpreted in terms of a limited GE for success, specific to the experimental procedure, but generalizing from task to task.

References

1. Buros, O. K. (Ed.) *The third mental measurements yearbook*. New Brunswick, N. J.: Rutgers Univer. Press, 1949.
2. Chapman, D. W., & Volkmann, J. A social determinant of the level of aspiration. *J. abnorm. soc. Psychol.*, 1939, **54**, 225–238.
3. Crandall, V. An investigation of the specificity of reinforcement of induced frustration. *J. soc. Psychol.*, 1955, **41**, 311–318.
4. Jessor, R. The generalization of expectancies. *J. abnorm. soc. Psychol.*, 1954, **49**, 196–200.
5. Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald Press, 1944. Pp. 333–378.
6. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
7. Rotter, J. B. Level of aspiration as a method of studying personality. *J. exp. Psychol.*, 1942, **31**, 410–422.
8. Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

(Received June 4, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

EASE OF ATTAINMENT OF CONCEPTS AS A FUNCTION OF RESPONSE DOMINANCE VARIANCE[1]

JONATHAN L. FREEDMAN AND SARNOFF A. MEDNICK

Harvard University

Underwood has suggested a method for the study of concept formation which assumes that the attainment of a concept calls for the perception of a relationship between concept instances. The perception of this relationship is chiefly dependent on two factors, the contiguity of the relevant associative responses to the concept instances, and the probability of the occurrence of the relevant associative responses to the concept instances. The latter is termed response dominance. The mean response dominance of the concept instances is termed dominance level (**3, 4**).

Underwood and Richardson have shown that the ease of attainment of a concept is directly related to dominance level (i.e., mean response dominance). Underwood and Richardson, while varying the *mean* response dominance of the concept instances, kept the *variance* within concepts low and constant across dominance levels (**5**). From Underwood's theoretical framework there is reason to suspect that the variance of the response dominances within concepts may also be an important variable. There are several ways the variance of the response dominance of concept instances could be manipulated. Either extreme may be changed or the extremes could be held constant and the middle values changed.

The clearest prediction from Underwood's theory comes in the case where variance is manipulated in the manner illustrated by the concepts in Table 1. It can be seen that the high variance (HV) concept is represented by one instance of very high response dominance and three instances of very low response dominance. The low variance (LV) concept is represented by four instances of approximately equal low response dominance. The dominance level is kept approximately equal across concepts. Under these conditions the high response dominance instance in the HV condition should elicit the correct concept response very early. This would place the correct response of the HV concept in contiguity with the relevant concept instances and should result in relatively easier attainment of the HV concept.

This experiment will compare the ease of attainment of concepts using HV and LV concept instances, manipulating variance in the manner described above. It is predicted that the HV concepts will be attained in fewer trials and with fewer errors than the LV concepts

Procedure

Materials.—The words used as stimuli were concrete nouns selected from a list of 213 nouns (**4**) for which Underwood and Richardson have ascertained the dominance level of various responses. Response dominance is defined as the percentage of the total number of responses given to a particular stimulus that fell in the given response category. For example, in Table 1, the 76 after "gnat" means that 76% of the responses given to "gnat" fell in the category "small." Concepts with a low mean dominance level were used in order to increase difficulty and facilitate construction of lists with greatest extremes of variance. The concepts "small" and "smelly" were chosen for two reasons: (*a*) there were enough instances for which these concepts were low dominant responses; (*b*) it was relatively easy to avoid instances to which both "small" and "smelly" were probable responses.

[1] This study was supported by a grant from the Laboratory of Social Relations.

TABLE 1

CONCEPT-LEARNING LISTS WITH RESPONSE DOMINANCE, MEAN DOMINANCE LEVEL, AND DEGREE OF VARIANCE INDICATED

Variance Level	List 1			List 2		
	Noun	Response Dominance	Concept and Mean Dominance	Noun	Response Dominance	Concept and Mean Dominance
High	gnat	76%	"small"	garlic	58%	"smelly"
	needle	9%		vinegar	14%	
	stone	7%		daffodil	12%	
	canary	5%	24%	coffee	12%	24%
Low	sauerkraut	24%	"smelly"	cradle	24%	"small"
	hospital	23%		closet	24%	
	tobacco	23%		rice	24%	
	gym	21%	23%	pin	22%	24%

As can be seen in Table 1, four groups were assembled, HV "smelly," HV "small," LV "smelly," and LV "small." This was necessary because of the possibility that the concepts might differ in difficulty or that concept difficulty may interact with variance level. List 1 consisted of HV "small" and LV "smelly" while List 2 contained the other two concepts. A buffer concept, "round" ("platter," "collar," "badge," "apple"), was used to make it more difficult for *S* to attain the concepts by elimination procedures.

An effort was made to have the one high dominant instance in the HV concepts as high as possible, although this was limited by the shortage of very low dominance instances which were necessary to equate dominance levels. As is shown in Table 1, the mean dominance levels of the concepts were kept nearly constant.

The instances were presented to *S* in three random orders. The same three orders were used for both lists, with positions occupied by HV instances in List 1 being occupied by LV instances in List 2. The positions of the buffer items were not changed between lists, although they were randomized in the three orders that were used.

Subjects.—The *S*s were 40 volunteers from Harvard and Radcliffe Colleges. The 18 women and 22 men were divided equally between the two lists.

Method.—The lists were presented at a 4-sec. rate on a Gerbrands-type memory drum. A 12-sec. interval occurred between presentations of the lists.

The following instructions, which were very similar to those used by Underwood and Richardson (**5**), were read to *S*:

"There are 12 nouns on the tape; four of these can be described by the same adjective, four by another adjective, and four by a third adjective. In other words, there are three groups of four words and all of the words in each group can be described by the same adjective. It is your task to discover what four nouns go in each group and what adjective can be used to describe each group.

"Each noun will be presented for 4 sec. and you must give a response to each one within the 4-sec. period. I will tell you if you are right or wrong. We will go through the list time after time and you are to give a response to each word every time it comes up even though you have been getting it right consistently.

"If you learn the list completely you will be responding with only three words because each of the words is correct for four of the nouns. These are three distinctly different words and are in no sense synonyms."

The *S* was made to understand that the type of response desired was an adjective that described sense impressions. If on the first trial *S* began giving the wrong type of response, *E* repeated that sensory impressions were required (this was necessary in only a few cases).

The test was continued to the criterion of one perfect trial, or terminated at 20 trials. A more complete discussion of the materials and procedure may be found elsewhere (**4, 5**).

RESULTS AND DISCUSSION

Nonparametric tests were used throughout to evaluate significance

level due to extreme irregularities in the distributions of trial and error scores and instances of heterogeneity of variance. This problem was also encountered by Underwood and Richardson (**5**).

Table 2 presents the data for the number of individuals who attained each concept on successive blocks of five trials. The criterion of attainment for a concept was one perfect trial. It can be seen that the HV concepts were solved significantly earlier and that significantly fewer individuals failed to solve them. Also the number of trials taken to attain each concept, as analyzed by the Wilcoxon matched pairs signed ranks test (**2**), was significantly less for the HV concepts ($T = 122$, $z = 3.16$, $P < .002$).

The total number of errors on each concept was counted for each individual. The median number of errors made on the HV and LV concepts were 19 and 40 respectively. Evaluated by the Wilcoxon matched pairs signed ranks test, this difference was significant ($T = 160$, $z = 3.36$, $P < .001$).

TABLE 2

NUMBER OF *Ss* ATTAINING CONCEPT ON SUCCESSIVE 5-TRIAL BLOCKS, AND MEDIAN NUMBER OF TRIALS TO ATTAINMENT

Trials	Variance Level		χ^{2*}	*P*
	LV	HV		
1–5	4	13	4.92	.05
6–10	12	15	.267	NS
11–15	9	6	.267	NS
16–20	4	2	.167	NS
Not attained	11	4	4.00	.05
Mdn.	12	8		

* Chi square for correlated proportions (**1**).

It was suggested previously that the HV concepts would be attained more easily because the one instance with high response dominance would elicit a correct response earlier in the test than would any of the four instances in the LV concepts. The median trial on which the first correct response was given was 4 for the HV concepts and 10 for the LV concepts. This difference was evaluated by the Wilcoxon matched pairs signed ranks test, and was found to be significant ($T = 84.5$, $z = 4.03$, $P = .00006$). This result supports the hypothesis regarding the manner in which HV operates. The high response dominance instance quickly places the correct concept response in contiguity with the other relevant concept instances. This tends to "polarize" responses to the relevant instances and results in rapid attainment of the HV concept.

It may be worth pointing out that while response dominance variance was cited as the independent variable in this study only one method of increasing variance was used; a single high dominance instance was combined with several low dominance instances. The consequent response polarization could have also been named as the independent variable. The advantage of the term "variance" is that it succinctly summarizes the experimental manipulation involved and contrasts it with the mean dominance level method. The polarization notion has the important advantage of suggesting the manner in which the variance variable operates. Because of this it may be more fruitful in specifying ways of manipulating variance.

SUMMARY

Forty *Ss* were presented with lists of 12 nouns and were instructed to discover into what three groups the nouns could be divided and what adjective could describe each group. The lists consisted of concepts of equal levels of dominance; the variance of the concept instances was manipulated. The high variance concepts

were attained more quickly and with fewer errors than the low variance concepts. An explanation of this is offered and is supported by the results.

References

1. Edwards, A. L. *Experimental design psychological research.* New York: Rinehart, 1950.
2. Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.
3. Underwood, B. J. An orientation for research on thinking. *Psychol. Rev.*, 1952, **59,** 209–220.
4. Underwood, B. J., & Richardson, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53,** 84–95.
5. Underwood, B. J., & Richardson, J. Verbal concept learning as a function of instructions and dominance level. *J. exp. Psychol.*, 1956, **51,** 229–238.

(Received June 7, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

ISOLATION EFFECTS: STIMULUS AND RESPONSE GENERALIZATION AS EXPLANATORY CONCEPTS [1, 2]

SLATER E. NEWMAN

North Carolina State College

AND

ELI SALTZ

Wayne State University

The present experiment on the von Restorff isolation effect has two purposes: (*a*) to examine the extent to which the concepts of stimulus and response generalization can be used to predict isolation phenomena in serial learning; and (*b*) to investigate the inconsistency between some of the results reported by Kimble and Dufort (**7**) and the results of other studies of isolation reported in the literature.

Von Restorff (**20**) found that an isolated term, such as a number in the midst of a list of nonsense syllables, was learned very quickly. Many explanations arising from entirely different theoretical frameworks have been proposed for the isolation effect. Thus, von Restorff used the gestalt concepts of figure and ground to predict her results, Green (**4**) refers to a "surprise" factor, while Smith and Stearns (**17**) have proposed that presence of the isolated item leads to reorganization of the serial list, the isolated term serving as an orienting device for the learner.

S-R psychologists have, in general, employed the notion of reduced intralist competition as an explanatory device (**1**; **2**; **8**, p. 384; **11**, p. 504; **13**). In addition, McGeoch and Irion (**8**), Osgood (**11**), and Underwood (**18**) suggest that the concepts of stimulus and response generalization may provide the basis for a theory of competition in verbal learning. Thus, the effects of isolation might be explained in terms of its effects on the tendencies for stimulus and response generalization to occur. The reasoning is this. The typical operational manipulation of stimulus generalization is to increase the similarity among *stimulus* members of a list. The typical operational manipulation of response generalization is to increase the similarity among *response* members of a list. In von Restorff-type experiments, the isolation of an item consists of inserting a dissimilar item into a list of similar items. Thus, the isolation manipulation can be expected to lead to a reduction in within-list stimulus and response similarity, and a consequent decreased likelihood that stimulus and response generalization will occur. On the

[1] This report is based on work done under ARDC Project No. 7709, Task No. 77304, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government. The opinions or conclusions expressed herein are those of the authors. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force or of the Air Research and Development Command.

[2] A paper, based on the results of this study, was read at the 1957 meeting of the Midwestern Psychological Association.

basis of generalization theory, then, it is possible to state a set of predictions concerning the effects of isolation on the learning of a serial list.

The isolated term as a response.—The most consistent finding of isolation studies is that the isolated term is well learned as a response. Can this be predicted from generalization theory? Consider two serial lists which are identical except for the item in the Xth serial position. In the first (isolated) list, the item in the Xth position is dissimilar from the remaining items, although these remaining items are similar to one another. In the second (homogeneous) list, the item in the Xth position is similar to all the other items in the list. First consider the operation of response generalization. Since the isolated term in the Xth position of the first list is dissimilar from the other items, the tendency for other responses to be inappropriately elicited in place of the isolated response should be less than the tendency for such items to be inappropriately elicited in place of the corresponding Xth item in the second (homogeneous) list. Thus, the isolated term should be learned more quickly as a response than its nonisolated counterpart. The concept of response generalization, then, can be used to predict the typical von Restorff effect in serial learning.

A second prediction can also be generated using response generalization. It was pointed out above that in the isolated list there should be a reduced tendency for other responses in the list to compete for elicitation with the isolated term. In a like manner, there should be a reduced tendency for the isolated term to be elicited inappropriately in place of any of the other responses in the list. This prediction is more crucial than the first, since it does not represent a post hoc rationalization of known data.

The isolated term as a stimulus.—Consider again the two hypothetical serial lists discussed above. Since the isolated term is dissimilar from other items in the list, there should be less tendency for the isolated item, as a stimulus, to elicit responses which are appropriate for other stimuli in the list than would be true for the corresponding Xth item in the homogeneous list. Thus, by using the concept of stimulus generalization, it would be predicted that isolation should reduce competition of responses to the isolated term as a stimulus. Consequently, isolation should facilitate the learning of a response to the isolated term as a stimulus.

Effect on learning the entire list.—Since the isolated term is dissimilar from the other items in the list, there should be a lessened tendency for stimulus and response generalization to occur throughout the isolated as contrasted with the nonisolated list. Thus, total errors during learning of the isolated list will be fewer than for the nonisolated list, even when the learning of the Xth items and their responses are not considered.

These predictions were evaluated using materials employed by Kimble and Dufort (**7**) in their Exp. III. Kimble and Dufort attempted to test the utility of conceptualizing "meaningfulness" as a dimension on which similarity could be manipulated. Thus, they attempted to manipulate isolation by placing a term low in Noble's (**10**) "meaningfulness" scale within a list of highly meaningful terms. Kimble and Dufort succeeded in obtaining the typical von Restorff effect: The isolated term was learned quickly as a response. However, contrary to the prediction made above, the response to the isolated term was learned more slowly than the response to the corresponding item in the homogeneous list. While the results of previous studies have not been conclusive on this issue (**5, 6, 17, 22**), the Kimble and Dufort results appear to be inconsistent with those studies which are methodologically most relevant (**6, 17**). Consequently, one of the purposes of the present study was to investigate the possibility that the Kimble and Dufort results were a peculiarity of the particular isolated items they employed.

It should be pointed out that evidence

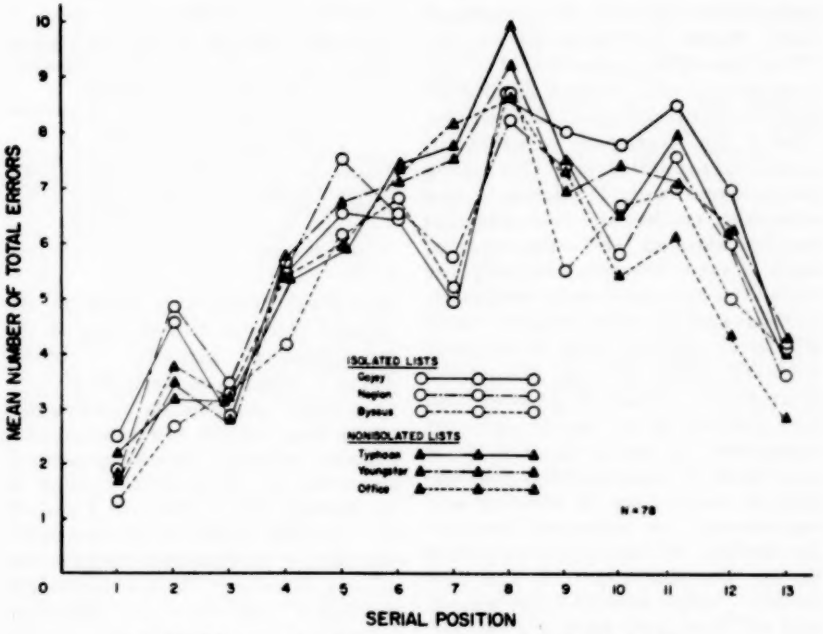

FIG. 1. Mean number of total errors as a function of serial position.

exists, from experiments reported in the literature, relevant to several of the predictions in this paper (**1, 3, 4, 5, 6, 7, 9, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22**). However, this evidence arises from several studies which used different procedures both for learning and for producing isolation. In the present experiment the predictions were tested within a single type of learning situation, using a single isolation technique.

METHOD

Learning materials.—The materials were similar to those used by Kimble and Dufort (**7**) in their serial learning study. Six 13-word lists, identical except for the seventh word, were constructed. The 12 words common to all lists were from the high end of Noble's meaningfulness scale (**10**) their *m* values ranging from 7.12 to 9.61. For the three nonisolated lists, the words in the seventh position also had high *m* values: youngster (6.75), typhoon (6.83), and office (7.95). For the isolated lists, the terms in the seventh position were paralogs of low *m* value; gojey (.99), neglan (1.04), and byssus (1.13). Each list was preceded by a double asterisk which served as the stimulus for the first term.

Subjects and procedure.—The *S*s were 78 airmen waiting to attend an Air Force technical school. Thirteen *S*s were randomly assigned to each group, a different list being learned by each group. Data from one *S* were discarded since he reported learning the list in narrative fashion.

After instructions describing the serial learning task were read to *S*, learning of the list was begun and proceeded to one errorless repetition of the list. The terms were exposed on a memory drum at a 2-sec. rate. Trials were separated by 12 sec.

RESULTS

Prediction 1.—The isolated term will occur more frequently as a correct response than will a nonisolated term occupying the same serial position. Figure 1 demonstrates that there are fewer errors at Position 7 for the

isolated lists than for the nonisolated lists. Table 1 presents an analysis of variance of the proportion of correct responses at this position divided by total correct responses at all positions. The *F* for treatments is significant at beyond the .001 level. This finding is consistent with Prediction 1 and with the results of other isolation studies showing that the isolated term is more effective as a response term than its nonisolated counterpart.

Prediction 2.—The isolated term will occur less frequently as an overt intrusion than will its nonisolated counterpart. Since the isolated item was learned more quickly than its nonisolated counterpart, and since also there is a complicated relationship between stage of learning and the tendency for intrusions to occur, an analysis of intrusions occurring in the first learning trial was performed. Application of Fisher's exact test to those data gave a *P* of .05, when a two-tailed test was employed. The results were, however, opposite in direction from those anticipated on the basis of Prediction 2. Thus, significantly more overt intrusions occurred involving the isolated terms than involving the nonisolated terms.

Prediction 3.—More correct responses will be made to an isolated term as a stimulus than to a nonisolated term occupying the same serial position. The serial position curves presented in Fig. 1 indicate

TABLE 1

ANALYSIS OF VARIANCE OF RATIO OF CORRECT RESPONSES AT POSITION 7 TO TOTAL CORRECT RESPONSES AT ALL POSITIONS

Source	*df*	*MS*	*F*
Treatments	1	120.78	21.32*
Groups/Treatments	4	6.68	1.18
Within groups	72	5.66	

* $P = .001$.

TABLE 2

ANALYSIS OF VARIANCE OF RATIO OF CORRECT RESPONSES AT POSITION 8 TO TOTAL CORRECT RESPONSES AT ALL POSITIONS

Source	*df*	*MS*	*F*
Treatments	1	11.85	3.04*
Groups/Treatments	4	6.60	1.69
Within groups	72	3.90	

* $.10 > P > .05$.

that fewer errors were made to the isolated terms as stimuli than to the corresponding nonisolated terms. Table 2 presents an analysis of variance based on these data, using the proportion, correct responses at this position divided by total correct responses at all positions. The *F* for Treatments is 3.04 with 1 and 72 *df*. On the basis of a two-tailed test, this *F* is significant between the .05 and .10 levels. Since a directional prediction was made, it appeared legitimate to evaluate the results using a one-tailed test; on the basis of a one-tailed test the *F* is significant at beyond the .05 level.

Prediction 4.—Total errors during learning of isolated lists will be fewer than for nonisolated lists, even when the isolated terms as stimuli and as responses are not considered.

Results of an analysis of variance of the number of errors (excluding those at Positions 7 and 8 which involve the isolated term as stimulus and as response), indicate that 59.74 errors were made in learning the isolated lists and 58.15 errors were made in learning the nonisolated lists.[3] The *F* for Treatments is less than 1.00 indicating that the difference is not significant. Thus, Prediction 4 fails

[3] Mean number of trials to learn the isolated lists, *including the isolated item*, was 12.9, to learn the nonisolated lists, 13.0. This lack of difference occurs despite the relatively rapid learning of the isolated term.

of verification. The F for Groups within Treatments was also less than 1.00.

Discussion

The results of the present experiment do not strongly support the utility of stimulus and response generalization as explanatory concepts for the effects of isolation. The stimulus generalization concept proved useful for predicting the facilitated learning of a response to an isolated stimulus; however, this effect was weak, and was statistically significant only when a one-tailed test was employed. The strongest isolation effect in the present experiment and in previous research was the tendency for an isolated item to be rapidly learned as a response. The writers found it necessary to employ the concept of response rather than stimulus generalization to make the essentially post hoc prediction of this phenomenon.

Although the concept of response generalization effectively handles the above results (i.e., the tendency for the isolated item to be rapidly learned as a response), it does not handle the intrusion data. Specifically, and contrary to prediction based on the response generalization concept, the data indicated that the isolated term occurred significantly more, rather than less, frequently as an intrusion error early in learning.

As was predicted earlier in this paper, a consequence of both stimulus and response generalization should be fewer errors during learning of isolated lists than of nonisolated lists. In the present study no such tendency was observed, the results being opposite in direction, although not significantly so, from those predicted. It may be argued that reduction of within-list similarity by one item may have only a very small effect on the learning of the entire list. It is interesting to note, however, that an examination of the data of three other studies also indicates poorer learning of the remainder of the isolated lists (**6, 7, 17**), although significance levels are not indicated in any of these studies. The writers are aware of no studies in which over-all facilitation has been reported.

One of the purposes of the present study was to investigate an inconsistency between results reported by Kimble and Dufort (**7**), and results reported by others (**6, 17**). Specifically, Kimble and Dufort reported that more errors were made to the isolated term as a stimulus than to the corresponding nonisolated term. The writers felt that this result might have been a function of the particular isolated term (i.e., "gojey") and the particular nonisolated term (i.e., "office") used by Kimble and Dufort. Consequently, in the present study three isolated and three nonisolated lists were constructed. "Neglan" and "byssus" were used as isolated items, in addition to "gojey"; while "youngster" and "typhoon" were used as nonisolated items in addition to "office." As was reported earlier in this paper, the isolated terms, as stimuli, elicited more correct responses than their nonisolated counterparts. However, fewer correct responses were elicited by "gojey" than by "office," which, as a stimulus, was the most effective of the three nonisolated terms (cf. Fig. 1). Thus, the Kimble and Dufort results on this point may be an artifact of the particular items they employed.

The present results suggest that the concepts of stimulus and response generalization may not be adequate to account for the effects noted. The alternatives, then, are these: (*a*) to identify additional processes, which, together with those of stimulus and of response generalization, can be used to explain the effects noted, or (*b*) to reject stimulus and response generalization as explanatory concepts, and to identify other processes which may handle more effectively the results obtained in the present study. At any rate, the view that all the effects of introducing an isolated term into a serial list are due to a reduction in intraserial interference may demand revision.

Summary

This study examined the extent to which the concepts of stimulus and response generalization can be used to predict isolation phenomena in serial learning. Four predictions based on those concepts were evaluated.

The learning materials were six 13-word lists, identical except for the seventh word. The 12 words common to each list had high meaningfulness. For the three nonisolated lists, the word in the seventh position was also of high meaningfulness; for the three isolated lists, they were low in meaningfulness. Each list was learned by a different group of 13 *Ss*. A serial anticipation procedure was employed.

As predicted, the isolated terms occurred correctly more frequently as responses than did the nonisolated terms and also tended, as stimuli, to elicit more correct responses than the nonisolated terms. Contrary to expectation, however, over-all learning of the isolated list was not facilitated; also the isolated terms occurred significantly more, rather than less, frequently as overt intrusions than the nonisolated terms. It is concluded that the concepts of stimulus and response generalization may not be adequate to account for the effects of isolation.

References

1. Buxton, C. E., & Newman, E. B. The forgetting of 'crowded' and 'isolated' materials. *J. exp. Psychol.*, 1940, **26,** 180–198.
2. Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.
3. Gibson, E. J. Intra-list generalization as a factor in verbal learning. *J. exp. Psychol.*, 1942, **30,** 185–200.
4. Green, R. T. Surprise as a factor in the von Restorff effect. *J. exp. Psychol.*, 1956, **52,** 340–344.
5. Jenkins, W. O., & Postman, L. Isolation and 'spread of effect' in serial learning. *Amer. J. Psychol.*, 1948, **61,** 214–221.
6. Jones, F. N., & Jones, M. H. Vividness as a factor in learning lists of nonsense syllables. *Amer. J. Psychol.*, 1942, **55,** 96–101.
7. Kimble, G. A., & Dufort, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50,** 361–368.
8. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
9. McGourty, M. C. Serial position effects in learning as a function of interfering associations. Unpublished M.A. thesis, State Univer. of Iowa, 1940.
10. Noble, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59,** 421–430.
11. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.
12. Pillsbury, W. B., & Raush, H. L. An extension of the Köhler-Restorff inhibition phenomenon. *Amer. J. Psychol.*, 1943, **56,** 293–298.
13. Postman, L., & Phillips, L. W. Studies in incidental learning: I. The effects of crowding and isolation. *J. exp. Psychol.*, 1954, **48,** 48–56.
14. Saul, E. V., & Osgood, C. E. Perceptual organization of materials as a factor influencing ease of learning and degree of retention. *J. exp. Psychol.*, 1950, **40,** 372–379.
15. Siegel, P. S. Structure effects within a memory series. *J. exp. Psychol.*, 1943, **33,** 311–316.
16. Smith, M. H., Jr. The influence of isolation on immediate memory. *Amer. J. Psychol.*, 1949, **62,** 405–411.
17. Smith, M. H., Jr., & Stearns, E. G. The influence of isolation on the learning of surrounding materials. *Amer. J. Psychol.*, 1949, **62,** 369–381.
18. Underwood, B. J. *Experimental psychology.* New York: Appleton-Century-Crofts, 1949.
19. Van Buskirk, W. L. An experimental study of vividness in learning and retention *J. exp. Psychol.*, 1932, **15,** 563–573.
20. von Restorff, H. Über die Wirkung von Bereichsbildungen im Spurenfeld. In W. Köhler and H. von Restorff, Analyse von Vorgängen im Spurenfeld. I. *Psychol. Forsch.*, 1933, **18,** 299–342.
21. Young, C. W., & Supa, M. Mnemic inhibition as a factor in the limitation of the memory span. *Amer. J. Psychol.*, 1941, **54,** 546–552.
22. Zirkle, G. A. Success and failure in serial learning: II. Isolation and the Thorndike effect. *J. exp. Psychol.*, 1946, **36,** 302–315.

(Received June 10, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

EFFECTS OF STRENGTH OF DRIVE ON LEARNING AND ON EXTINCTION [1]

HERBERT BARRY, III

Yale University

Animals performing under low drive may readily be compared with other animals performing under high drive. A difference in performance under different drive strengths does not distinguish between two alternative possibilities: (*a*) performance may differ only as long as drive is different; or (*b*) the groups may become different in a more permanent respect, such as habit strength, so that performance continues to differ even after drive is equalized.

A subsequent test under equalized drive strength may distinguish between these alternative possibilities. An appropriate method for equalizing drive strength is to shift half the high-drive *Ss* to low drive (HL) while half remain under high drive (HH), and similarly to shift half the low-drive *Ss* to high drive (LH) while half remain under low drive (LL). This results in a 2 × 2 factorial design, with four experimental groups. The groups which differ in drive in the first stage are equalized in the second stage, and the groups which differ in drive in the second stage are equalized in the first stage.

The present experiment expands this design into three stages—training, early extinction, and later extinction—yielding a 2 × 2 × 2 design with eight groups. The groups are listed in the first column of Table 1, where they are designated by their drive (H or L) in the three successive stages (HHH, etc.). The first two stages provide a comparison between groups trained under different drive strengths and equalized in drive during early extinction. The addition of the third stage enables a comparison to be made between groups which differ in drive strength during extinction trials and are equalized in drive during further extinction.

Contradictory results have been reported in different experiments which have attempted to separate the effects on performance of previous and current drive strength. Among studies which have used the 2 × 2 design to measure effects on running speed, performance in the second stage was predominantly influenced sometimes by current drive (**7, 12**), and sometimes by previous drive (**4**).

Heathers and Arakelian (**6**) used a 2 × 2 design in a bar-pressing situation to test whether drive strength during early extinction affected performance after drive was equalized. They found that the number of bar presses made during later extinction was not reliably influenced by drive during early extinction. However, since they trained all *Ss* under high drive the interpretation of this result requires the questionable assumption that drive during training does not alter the effects of drive during extinction. Furthermore, a long time

[1] This report presents part of a dissertation submitted to the Department of Psychology, Yale University, in partial fulfillment of the requirements for the Ph.D. degree. The project was helped by Neal E. Miller's research grant, M647, from the National Institute of Mental Health of the National Institutes of Health, Public Health Service, Department of Health, Education, and Welfare, Bethesda, Md., and by Frank A. Logan's research grant from the National Science Foundation. The author is grateful for the advice and encouragement given by Miller and Logan and in addition by Fred D. Sheffield and George F. Mahl.

interval was imposed between early and later extinction, so that recovery from the effects of early extinction might have occurred.

In the above studies, and in others cited by Barry (2), the change in drive strength at the time of testing was accomplished by a change in feeding or testing schedule. If *S*s are tested every day, as in previous studies, a change in drive requires changing either the time of feeding or the time of testing. The present two-day schedule, which is described below, was designed to avoid confounding changes in drive with either of these two factors.

Method

Subjects.—The *S*s were 64 naive male albino rats of Sprague-Dawley strain, obtained from the Holtzman Co., Madison, Wis. They were 105–145 days old at the beginning of the experiment.

Schedule.—All *S*s were run on a two-day schedule, which was repeated throughout the course of the experiment. This schedule involved three successive procedures. (*a*) The first feeding, which consisted of Purina lab chow for 1 hr., followed immediately by ½ hr. of the food pellets used in testing. Thus, *S*s were thoroughly satiated on the reward pellets. (*b*) Testing, which occurred after a deprivation time, measured from the end of the first feeding, of either 2½ hr. (low drive) or 26½ hr. (high drive). The low-drive and high-drive tests were at the same time of day, since they differed by exactly one day in deprivation time after the first feeding, which in turn was at the same time of day for all *S*s. (*c*) The second feeding, given ½ hr. after the high-drive test or 24½ hr. after the low-drive test. It consisted of 20 gm. of Purina chow. The first feeding was given again 20½ hr. after the second feeding (48 hr. after the prior first feeding).

Under these procedures, *S*s which were tested under different drive strengths did not differ in feeding schedule or time of day of the test. A necessary feature for these conditions was the one-day difference in deprivation time between low-drive and high-drive tests. The one-day difference in deprivation time was made possible by the two-day schedule, which provided an interval of more than one day between the first and second feeding.

A change in drive strength was accomplished by changing the time of the test by one day in the two-day schedule, so that *S*s were tested on their previous non-test day. This procedure changed the number of hours of deprivation after the first feeding without changing the feeding schedule or time of day of the test. The time interval for this change was counterbalanced by testing half the *S*s one day and half the *S*s three days after the previous test, instead of at the usual two-day interval. No significant difference was found in subsequent performance between *S*s tested one day and *S*s tested three days after the last test under the previous drive strength.

Apparatus.—The apparatus was a straight alley 4 ft. long, described in detail elsewhere (10). Running time was measured by Standard Electric timers reading to 1/100 sec. Different timers were connected successively to a microswitch which was released when *E* opened the starting door of the alley, and to four photoelectric tubes spaced 12 in. apart, the first one 4 in. from the starting door and the last one 2 in. from the food cup. The *S*s were housed in individual cages in a separate room and were brought into the laboratory in individual wooden boxes 8 in. square, each with a hole in one side for a water tube. Immediately before and after each trial *S*s were placed in a wooden box 18 in. square for approximately 2 min.

Experimental design.—The *S*s were tested every other day, during three stages: (*a*) Training for 21 testing days with 1 trial on each of the first two days, 2 trials on Day 3, 4 on Days 4–13 and 1 on Days 14–21, for a total of 52 trials. (*b*) Early extinction for six testing days, at one trial per day. (*c*) Later extinction for 12 testing days, at one trial per day. When two or more trials were given on the same day, they were spaced 20 min. apart. Half the *S*s were trained under low drive and half under high drive. At the start of early extinction, and again at the start of later extinction, half the *S*s in each group were shifted to the other drive and half continued under the same drive as before. The result was a $2 \times 2 \times 2$ design with the eight experimental groups which are shown in Table 1. There were 8 *S*s in each group.

Testing procedures.—On each training trial *E* placed *S* in the starting compartment of the alley and 15 sec. later raised the starting door. If *S* turned back after it had entered the alley, the door of the starting compartment was lowered. When *S* reached the last phototube, two food pellets (.045 gm. each, purchased from the P. J. Noyes Co., Lancaster, N. H.) were automatically released with an audible click down a chute into the food cup. The *E* lowered the goal-box door immediately. If *S* ate both pellets within 6 sec., it was lifted out of the goal box 8 sec. after the delivery of food; otherwise

TABLE 1

AVERAGE RUNNING SPEEDS OF EACH GROUP DURING THE THREE STAGES OF THE EXPERIMENT

Group	Training	Early Extinction		Later Extinction			
	45–52	1–3	4–6	1–3	4–6	7–9	10–12
HHH	2.28	2.27	2.01	1.50	1.10	.82	.80
HHL	2.33	2.40	1.91	1.33	.89	.81	.76
HLH	2.29	1.77	1.26	1.11	1.10	.75	.76
HLL	2.28	1.79	1.26	1.01	.97	.70	.63
LHH	1.56	1.86	1.36	1.30	1.04	.94	.76
LHL	1.54	1.70	1.48	1.01	.78	.61	.60
LLH	1.50	1.57	1.27	1.43	1.33	1.26	1.05
LLL	1.54	1.61	1.18	.85	.64	.50	.50

Note.—Average running speeds are shown in ft./sec.

it remained in the goal box for 2 sec. after eating both pellets. During extinction trials the delivery mechanism clicked but did not deliver pellets, and *S* was removed from the goal box 8 sec. after reaching the last phototube.

Statistical procedures.—The time scores were converted into reciprocals, which gave a measure of speed with an approximately normal distribution of scores. The results reported are the reciprocals for the 4-ft. distance between the first and last phototubes. The results were tested by analysis of variance, separately for the average of the last 8 training trials, the 6 early extinction trials, the first 3, first 6, and last 6 later extinction trials. The variables tested were drive in the current stage and in all previous stages, and also the four replications and interactions of the experimental variables with each other and with the replications.

An analysis of covariance was also performed for the results during extinction to obtain a more sensitive test by taking into account the individual differences within groups which remained constant in the successive stages. Each *S*'s average speed was converted into a standard score for the last 8 training trials and separately for the last 3 trials of early extinction, in relation to the mean and standard deviation for all the *S*s in the same experimental condition. An analysis of covariance for early extinction was calculated by combining the standard scores during training with the analysis of variance for early extinction. Analyses of covariance for later extinction were calculated by combining the standard scores during early extinction with the analyses of variance for later extinction. The results were in the same direction for the analyses of variance and of covariance, but were more reliable for the latter. The statistical tables are shown in Barry (**2**).

RESULTS AND DISCUSSION

Table 1 shows the average speeds of each of the eight experimental groups at the end of training and throughout extinction. The different experimental effects may be determined by comparing appropriate combinations of groups.

Effects of drive during training.—Figure 1 compares the four groups under high drive during training and equalized throughout extinction (HHH, HHL, HLH, and HLL) with the four groups under low drive during training and equalized throughout extinction (LHH, LHL, LLH, and LLL). The average speeds during training under low and high drive are shown in the left-hand portion of Fig. 1. On days in which two or more trials were given, the average speed is shown only for the first trial of the day.[2] Both the low- and high-

[2] Averaging together the days in which four trials were given, and comparing Trial 1 with Trial 4, running speed was faster on Trial 4, both under high drive ($P < .01$) and under low drive ($P < .001$). This difference was reliably greater for the low- than high-drive *S*s ($P < .05$). On the other hand, in starting speed there was a decline from Trial 1 to Trial 4 both under high drive ($P < .01$) and under low drive ($P < .05$). The averages for Trials 2 and 3 were intermediate between Trials 1 and 4, both in running and starting speeds.

drive *Ss* learned to run to the food. The two curves start at the same low level of speed and diverge during the course of training. For the last eight training trials (at one trial per testing day) the average speed was 2.30 ft./sec. under high drive and 1.53 ft./sec. under low drive; the difference was highly reliable ($F = 32.2$, $P < .001$; $1, 56 df$). There was no significant change in speed during these eight trials; thus it may be concluded that the low- and high-drive *Ss* had reached, or were close to, different asymptotes of performance.

The right-hand portion of Fig. 1 shows that the difference in performance persisted after drive was equalized during extinction. The average speed for the six trials of early extinction was 1.84 ft./sec. after high drive and 1.50 ft./sec. after low drive during training. This difference was highly reliable ($F = 20.8$, $P < .001$; $1, 47 df$). It may be seen that the curves converged rapidly during later extinction.

Tolman (**15**) has suggested that *Ss* which perform differently under different motivational conditions may be equal in learning, so that the difference in performance may persist only as long as the difference in motivation is maintained. According to Hull's (**8**) theory, drive strength affects vigor of performance but does not affect habit strength, so that when drive is equalized performance should be immediately equalized. The present finding is contrary to these predictions.

If *Ss* trained under low and high drive learn qualitatively different responses as a result of running at different speeds, the difference in performance may be expected to persist after drive is equalized. This assumption was made by Campbell and Kraeling (**4**) and is a plausible explanation for their findings. They gave only 14 training trials, at the end of which the average speed was more than

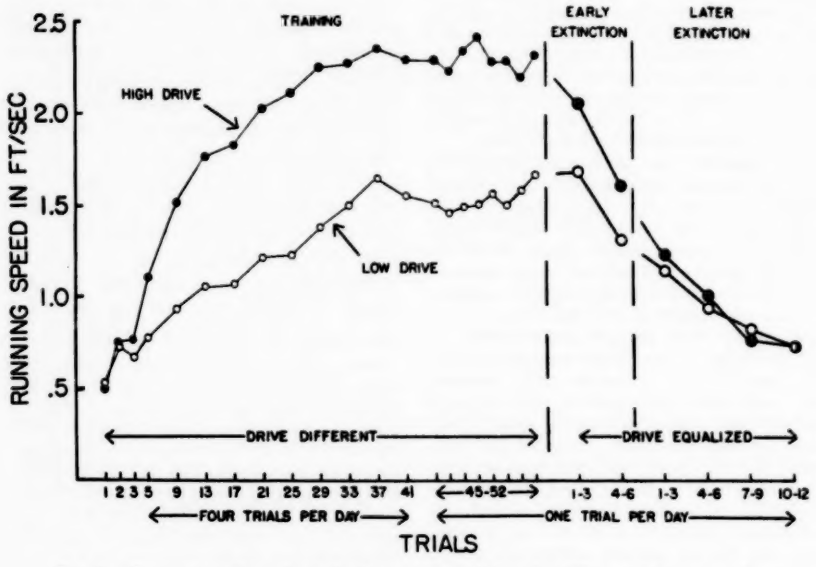

FIG. 1. Comparison throughout the three stages of *Ss* under different drive strengths during training and equalized in subsequent drive.

three times as fast for the high-drive as for the low-drive *S*s. According to Logan's (**9**) micromolar assumption, any differences in vigor of response may be considered as different responses rather than as different strengths of the same response.

According to alternative theories, the effect of drive during training persisted not because of the difference in performance during training but because of a difference in (*a*) strength of the anticipatory goal response that was learned; (*b*) habit strength; (*c*) amount of conditioned excitement, according to a possible interpretation of Sheffield's (**14**) theory; or (*d*) conditioning of the drive strength itself, as proposed by Hull's (**8**) theory of secondary drive.

Effects of drive during early extinction.—Figure 2 compares the four groups under high drive during early extinction (HHH, HHL, LHH, and LHL) with the four groups under low drive during early extinction (HLH, HLL, LLH, and LLL). During training these *S*s differed only in their future conditions, so that the running speeds were closely matched. During early extinction the average speed was 1.88 ft./sec. under high drive and 1.46 ft./sec. under low drive. This difference was highly reliable ($F = 29.0$, $P < .001$; 1, 47 *df*). During later extinction these *S*s were equalized in drive. In the first three trials of later extinction the average speed was 1.29 ft./sec. after high drive and 1.10 ft./sec. after low drive during early extinction. This difference was reliable ($F = 5.19$, $P < .05$; 1, 31 *df*). The effect of drive during early extinction on running speed disappeared so rapidly in subsequent trials of later extinction that there was a reliable change in this effect from the first three to the next three trials of later extinction ($F = 4.76$, $P < .05$; 1, 32 *df*).

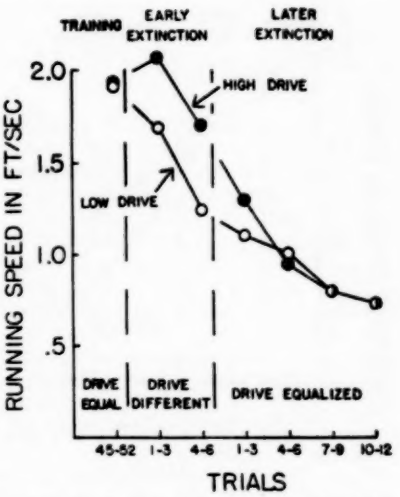

FIG. 2. Comparison at the end of training and during extinction of *S*s under different drive strengths during early extinction and equalized in drive in the preceding and subsequent stages.

The results suggest that extinction was less effective under high than under low drive. The reverse finding would be expected if one assumed that the effectiveness of extinction is in proportion to the vigor of the nonrewarded response or to the intensity of the frustrated motive. Hull (**8**) has emphasized vigor and Amsel (**1**) has emphasized frustration in their theories of extinction, so that additional principles would be necessary to make these theories consistent with the present results.

The finding of less effective extinction under high drive than under low drive might be expected from the assumption that secondary reinforcement is stronger under high drive, or from Sheffield's (**14**) theory that more conditioned excitement becomes associated with the alley under high drive. This finding might also be explained by the micromolar assumption that differences in running speed during early extinction are different responses, if the running speed which occurred was learned rather than inhibited during these nonrewarded trials. According to Guthrie's (**5**) theory,

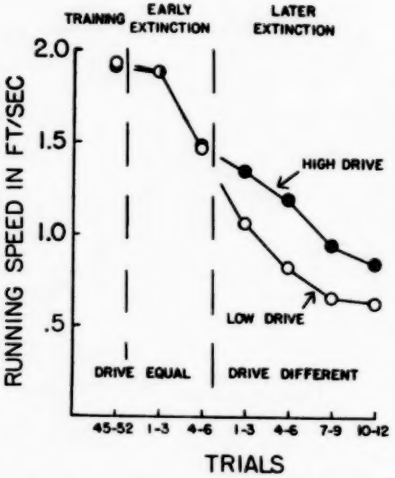

FIG. 3. Comparison at the end of training and during extinction of *S*s under different drive strengths during later extinction and equalized in drive in both preceding stages.

extinction is a process of learning alternative responses. The *S*s under low drive, which make the alternative responses more often, have more opportunity to learn them. Hull (**8**) viewed extinction as a process of learning the inhibitory response of stopping. If the slower running by *S*s under low drive is interpreted as a greater frequency of stopping, these *S*s would be expected to learn the inhibitory response more strongly.

Extinction might be regarded not as the absence of all reward but as a decrease in the amount of reward. This was the explanation given for a previous finding by Barry and Logan (**3**) that hungry rats learned to run to the end of the alley without food reward. If the micromolar assumption is made, high-drive *S*s during extinction would learn to run faster than low-drive *S*s because of being reinforced somewhat for running faster. This theory assumes that the same responses are learned in extinction as in training, whereas Hull's and Guthrie's theories assume that extinction is a process of learning responses different from those learned during training.

Effects of drive during later extinction.—Figure 3 compares the four groups under high drive during later extinction (HHH, HLH, LHH, and LLH) with the four groups under low drive during this stage (HHL, HLL, LHL, and LLL). It may be seen that the two curves were closely matched in running speed in both prior stages. During the first six trials of later extinction the average running speed was 1.24 ft./sec. under high drive and .94 ft./sec. under low drive; this difference was highly reliable ($F = 18.6$, $P < .001$; 1, 31 *df*). During the last six trials the average speed was .89 under high drive and .64 under low drive; this difference continued to be reliable ($F = 7.70$, $P < .01$; 1, 31 *df*). Although the conditions at the end of later extinction eliminated the persisting effects of drive differences during previous stages (Fig. 1 and 2), the performance effect of current drive remained.

According to Hull's (**8**) theory, high drive increases the performance of any learned responses. In the present experiment, the learned response during both training and extinction was running down the alley. The effect of the inhibitory response learned in extinction was to weaken but not eliminate or replace the running response. Therefore high drive would be expected to increase running speed not only during training but also in both stages of extinction. Guthrie's (**5**) theory regards extinction as the learning of new responses which replace the prior responses. Since high drive facilitated learning of the prior responses, one might expect it to facilitate the new learning involved in extinction. This prediction is contrary to the present finding that *S*s under high drive ran faster than *S*s under low drive throughout extinction as well as during training. However, Guthrie might conceivably make the new assumption that the strength of the hunger drive would affect responses associated

with food and not responses learned in the absence of food, during extinction.

Effects of drive change with drive strength equalized.—In both stages of extinction *Ss* under a different drive from a previous stage may be compared with *Ss* under the same drive in both stages. The *Ss* changed from low to high and from high to low drive are averaged together and compared with the average of *Ss* remaining under the same drive in both stages.

Figure 4 shows that during early extinction performance was impaired by a change to a different drive from training. The average running speed was 1.78 ft./sec. for *Ss* under the same drive as during training (HHH, HHL, LLH, and LLL) and 1.56 ft./sec. for *Ss* under a different drive in the two stages (HLH, HLL, LHH, and LHL). The difference was reliable ($F = 9.41$, $P < .01$; 1, 47 *df*).

Drive strength has been considered by Hull (**8**) as providing a distinctive stimulus. Pavlov (**11**) reported that the introduction of a novel stimulus inhibited a conditioned response. The conditioned response, which has been learned to a particular conditioned stimulus, may be regarded as weakened by stimulus generalization. Yamaguchi (**16**), who measured latency of bar pressing, found that *Ss* tested under a different drive strength from training responded more slowly the larger the difference between the two drives. He explained this result from the gradient of generalization for drive stimulus.

Figure 4 shows that an opposite effect was found during later extinction: performance was superior if drive was different from training. During the first six trials of later extinction, the average speed was 1.16 ft./sec. for *Ss* under different drive from training (HHL, HLL, LHH, and LLH) and 1.01 ft./sec. for *Ss* under the same drive as during training (HHH, HLH, LHL, and LLL). The difference was reliable ($F = 6.58$, $P < .05$; 1, 31 *df*).

Pavlov (**11**) reported that, although a novel stimulus inhibits a conditioned response, the novel stimulus when introduced after extinction trials typically "disinhibits" the conditioned response. He concluded that the inhibitory response is more unstable than the excitatory response, so that a novel stimulus removes a larger proportion of inhibitory than excitatory response. Saltz (**13**) gives a similar explanation by describing the gradient of stimulus generalization as steeper for inhibitory responses than for excitatory responses.

If drive strength provides a distinctive stimulus, disinhibition is a possible explanation for the finding that performance in later extinction was superior

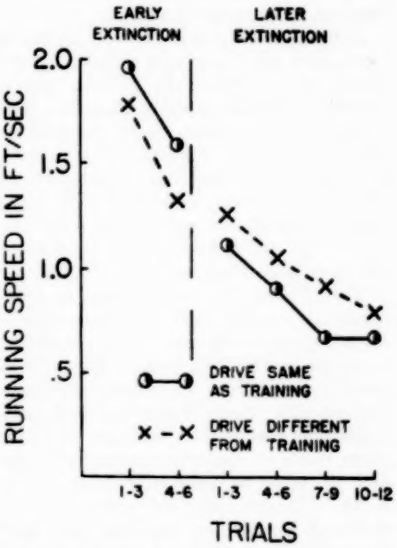

FIG. 4. Comparison of *Ss* currently under a different drive strength than during training with *Ss* currently under the same drive as during training, as shown during early and later extinction. The comparison represents different combinations of groups in early and later extinction.

if drive was different from training. Since the majority of previous trials were during training, a different drive during later extinction from training, regardless of the drive during early extinction, might have had the same effect as the introduction of a novel stimulus. The effect of introducing a novel drive strength after training and early extinction under the same drive may be shown by the comparison of Groups HHL and LLH with Groups HHH and LLL. This corresponds to Pavlov's procedure for demonstrating disinhibition. During the first six trials of later extinction, the *S*s which were changed at that point to a novel drive ran reliably faster ($F = 5.66$, $P < .05$; 1, 15 df). This is one of several separate comparisons, shown by Barry (2), which contribute to the finding of faster speed during later extinction by *S*s under different drive from training. The other comparisons show nonsignificant differences in the same direction.

General features of extinction.—During extinction the average running speed decreased progressively. There was a reliable decrease even within the last stage, from the first six to the last six trials of later extinction ($t = 6.77$, $P < .001$, with 64 df). Therefore all the experimental results which have been reported for extinction were obtained in a situation of continuous decrease in running speed.

The individual differences in performance among *S*s in the same experimental group remained highly consistent from one stage to the next. The within-group correlation coefficient between the last eight trials of training and early extinction was +.85; between the last three trials of early extinction and the first six trials of later extinction it was +.88. These high correlations were obtained in spite of the fact that half the groups were changed in drive at the end of training and early extinction, and over-all speed changed rapidly during extinction.

The results have been presented in terms of speed for the entire 4-ft. alley. In extinction the decrease in speed was greater, the closer to the empty food cup it was measured. Combining all *S*s, the speeds in the last (18th) extinction trial in the starting section and in Sections 1, 2, 3, and 4 were, respectively, 95%, 79%, 64%, 34%, and 29% of the last training trial. The detailed results for each 1-ft. section and for starting into the alley are presented by Barry (2). Most of the experimental effects were found in the same direction for each of these measures.

Summary

A procedure was used which determined whether differences in strength of drive (*a*) affected current performance and/or (*b*) affected subsequent performance after drive was equalized. This design was used to investigate the effects of differences in drive during learning and extinction.

The procedure consisted of a $2 \times 2 \times 2$ factorial design with three stages (training, early extinction, and later extinction) and two strengths of hunger drive (low and high) in each stage. Thus there were eight experimental groups, which represented all possible sequences of the two drive strengths in the three stages. Albino rats were given spaced trials in a straight alley; two food pellets were delivered when *S* reached the end of the alley in each training trial but not during extinction.

The *S*s under high drive ran faster than *S*s under low drive in all three stages of the experiment. Thus with previous drive equalized, performance was increased by high drive even at a late stage of extinction.

The *S*s under high drive during training continued to run faster than *S*s under low drive for several extinction trials after drive was equalized. This result is contrary to Hull's prediction that *S*s trained under different drive strengths differ in performance only as long as drive remains different. The *S*s under high drive during early extinction continued to run faster than *S*s under low drive for a few trials of later extinction, after drive was equalized. This result shows that extinction is not explainable as a positive function of the vigor of the nonrewarded responses or the intensity of the frustrated motive.

Effects of changes in drive strength, with low and high drive averaged together, were also found. During early extinction a different drive from training resulted in slower running; during later extinction a different drive from training resulted in faster running. These results may

be interpreted as effects of change in drive stimulus.

References

1. Amsel, A., & Hancock, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53,** 126–131.
2. Barry, H., III. Effects of strength of drive on learning and on extinction. Unpublished Ph.D. dissertation, Yale Univer., 1957.
3. Barry, H., III, & Logan, F. A. Latent learning in a straight alley. *Psychol. Rep.*, 1957, **3,** 88.
4. Campbell, B. A., & Kraeling, D. Response strength as a function of drive level during training and extinction. *J. comp. physiol. Psychol.*, 1954, **47,** 101–103.
5. Guthrie, E. R. *The psychology of learning.* (Rev. ed.) New York: Harper, 1952.
6. Heathers, G. L., & Arakelian, P. The relation between strength of drive and rate of extinction of a bar-pressing reaction in the rat. *J. gen. Psychol.*, 1941, **24,** 243–258.
7. Hillman, B., Hunter, W. S., & Kimble, G. A. The effect of drive level on the maze performance of the white rat. *J. comp. physiol. Psychol.*, 1953, **46,** 87–89.
8. Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.
9. Logan, F. A. A micromolar approach to behavior theory. *Psychol. Rev.*, 1956, **63,** 63–73.
10. Logan, F. A., Beier, E. M., & Ellis, R. A. Effect of varied reinforcement on speed of locomotion. *J. exp. Psychol.*, 1955, **49,** 260–266.
11. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep). London: Oxford Univer. Press, 1927.
12. Ramond, C. K. The effect of a shift in time of deprivation upon performance in instrumental learning. *Proc. Iowa Acad. Sci.*, 1953, **60,** 540–545.
13. Saltz, E. A single theory for reminiscence, act regression, and other phenomena. *Psychol. Rev.*, 1953, **60,** 159–171.
14. Sheffield, F. D., Roby, T. B., & Campbell, B. A. Drive reduction versus consummatory behavior as determinants of reinforcement. *J. comp. physiol. Psychol.*, 1954, **47,** 349–354.
15. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.
16. Yamaguchi, H. G. Gradients of drive stimulus (S_D) intensity generalization. *J. exp. Psychol.*, 1952, **43,** 298–304.

(Received June 10, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

HALF-LOUDNESS JUDGMENTS WITHOUT PRIOR STIMULUS CONTEXT [1]

W. R. GARNER

The Johns Hopkins University

For purposes of constructing loudness scales (or other sensory scales) there are certain apparent advantages to the use of experimental procedures which require direct numerical estimates by *O* of the loudness of a given sound, or the loudness relations between two different sounds. These techniques have a face validity for scaling purposes which is appealing to many psychologists, and the result is that a considerable amount of research on loudness scaling has been done with these techniques. Unfortunately, however, these many researches have made it abundantly clear that different experimental procedures give quite different types of loudness function (**2, 4**), a fact which is disquieting, to say the least.

Garner (**1**) recently showed that when a method of constants is used, in which *O* judges whether the second of a pair of tones is more or less than half as loud as the first, the context of comparison stimuli presented to *O* can vary his "half-loudness" value by as much as 20 db. One obvious conclusion from these results is that it would be desirable to get judgments from *O*s before any context has been presented to them. Stevens and Poulton (**5**) have presented some data where the first adjustments of *O*s are used, or where the first direct magnitude estimate is used. Even these methods have some disadvantages, however. The method of adjustment provides some context since the attenuator to be used by *O* must be preset to some value. And both the method of adjustment and the method of direct magnitude estimation pose problems concerning the proper averaging techniques, since the distributions of judgments are rarely symmetrical. These problems are lessened to some extent with the method of constants, since only one exposure of *O* to the stimuli need be made, and since the final value obtained is simply an indifference point, i.e., the point at which the number of greater judgments equals the number of lesser judgments. Thus, no numerical averaging is necessary.

The purpose of the present paper is to present the results of an experiment in which *O*s were required to state whether the second of two tones was more or less than half as loud as the first. In order to avoid any context effects, it was required that only one judgment be used per *O*, and that this judgment be obtained without any preliminary practice. This last requirement caused some difficulties in the experiment.

METHOD

Observers.—All of the *O*s were male college students who were taking the elementary course in psychology, and who were required to serve as *O*s as part of the course. The total number of *O*s used was 135, and none had previously been used in auditory experiments. All of them had an absolute threshold for a tone of 1000 cps of 15 db. SPL (db. re .0002 microbar)

[1] This research was done under Contract N5-ori-166, Task Order 1, between the Office of Naval Research and The Johns Hopkins University. This is report No. 166-I-209, Project Designation No. NR 145-089, under that contract. The author is indebted to Elliot M. Cramer for assistance in the collection of the data.

or less, but this threshold was determined after the experiment in order to avoid any prior contextual experience, and the data were discarded in the few cases where the threshold was too high.

Procedure.—The general procedure is best described by the written instructions to *O*.

"The purpose of this experiment is simply to find out how different two tones must be in loudness for one of them to sound half as loud as the other. We are not concerned with the physical differences in intensity between the two tones, but only in what *sounds* half as loud to you. Thus, there is no question of your being right or wrong in your judgments, since we only want to know how these tones sound to you—and what you say is always right.

"You will be listening to a series of pairs of tones, one pair occurring every 7 sec. The first tone of the pair lasts for 1 sec., and always has the same loudness each time it occurs. Then there is a brief silent period before the second tone comes on and also lasts for a second. The second tone of the pair will vary in loudness from one time to another. Your task is to decide whether the second tone is more or less than half as loud as the first tone. If the second tone seems to you more than half as loud, you put a plus (+) on your record sheet. If it sounds less than half as loud, you put a zero (0) on your record sheet. You must always decide whether it is more or less than half as loud, and cannot say that it is exactly half as loud.

"There is a microphone in the room so that we can hear you when you talk. After each pair of tones, say out loud whether it seems more or less than half as loud to you—just before or at the same time you mark it on the record sheet. This way we will have a double check on your judgments."

There were additional instructions having to do with later judgments, since the data for the present experiment were collected as part of another experiment. After reading the instructions, *O* was seated in a soundproof room and a headset wired for monaural listening was fitted onto his head. Then before the first pair of stimuli was presented, *O* was asked to state verbally what he was required to do. These somewhat elaborate instructions and precautions were taken to be sure of getting a legitimate response to the very first stimulus pair, and *O* at no time prior to the first stimulus presentation was allowed to hear what he was asked to judge. Even when *O* entered the experimental room, the stimulus apparatus was turned off so that he would not inadvertently hear any tones.

All tones had a frequency of 1000 cps, and both the standard and comparison tones had durations of 1 sec., with a 1-sec. silent period between the two tones. The tones were filtered through a narrow band-pass filter to eliminate transients. All timing was done electronically with equipment in a room adjacent to that of *O*, who was alone during the experiment.

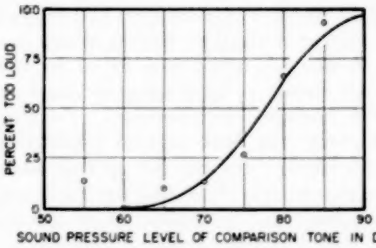

FIG. 1. The percentage of judgments that the intensity of the comparison tone indicated on the abscissa was more than half as loud as the standard tone (90 db.). Each plotted point represents the first judgments of 15 different *O*s, except those at 65 db. and 75 db., where 30 *O*s were used. The smooth curve is a least-squares solution based on the four highest intensities only.

Experimental conditions.—The only difference between the stimuli heard and judged by the different *O*s was in the intensity of the second tone of the pair. The standard tone had an intensity of 90 db. SPL for all *O*s, and the comparison tone had an intensity of either 85, 80, 75, 70, 65, 60, or 55 db. SPL. At intensities of 65 and 75 db., 30 different *O*s were used; and 15 *O*s were used at all other intensities. The larger number of *O*s at two intensities was due to the fact that the same *O*s were used in another experiment where that number was necessary for counterbalancing purposes.

RESULTS

The only data of this experiment are the number of *O*s who stated that each intensity of the comparison tone was more or less than half as loud as the standard tone, and these data are presented (in percentage form) in Fig. 1. At an intensity of 85 db., 14 out of 15 *O*s stated that the second tone was more than half as loud, and these numbers decreased down to 60 db., where none of the 15 *O*s stated that the second tone was more than half as loud. At 55 db., 2 out

of 15 *O*s again stated that the second tone was more than half as loud.

A normal ogive was fitted to the data using a least-squares solution with Muller-Urban weights. For this fit, only the four highest intensities were used, since the use of the lower values would have unduly flattened the function, and the curve drawn in Fig. 1 is this fitted normal ogive. This curve gives a value for half loudness of 77.6 db., or an attenuation from the standard of 12.4 db. This value is relatively unaffected by the number of points used in the least-squares solution, since the inclusion of more points has more effect on the slope of the function than on its 50% point.

The standard error of this half-loudness value is 1.37 db., using Culler's approximation (**3,** p. 132). This standard error gives a confidence interval at the .05 level of 75.4 db. to 79.9 db.

Discussion

When this experiment was undertaken, there were several questions about the outcome that seemed in doubt. One of these questions was whether a reasonable psychophysical function could be obtained at all when only one judgment per *O* was used. The previous research with the method of constants (**1**) had shown such a complete effect of context on the results obtained that it seemed possible that first judgments would be essentially random. The results of the present experiment make it clear that first judgments are meaningfully related to the intensity of the comparison tone, and that the context effects previously shown must in some way be superimposed on the initial judgment. It seems most likely that, while a first-judgment function can be obtained, these judgments are made with such a low level of confidence that the context provided in the experiment can easily influence the later judgments.

Another question, of course, had to do with the actual half-loudness value obtained. Stevens (**4**) has summarized many of the investigations on half-loudness values, and the variability of these results is the one outstanding feature of them. In the intensity range used in this experiment, attenuations necessary for half loudness from under 6 to over 20 db. have been reported. Stevens himself has suggested that an attenuation value of 10 db. is most valid, although when he and Poulton (**5**) used estimations and adjustments on first trials only, they obtained values larger than this for most of their conditions. It has been recognized that first judgments, because of the severe context effect, are more valid than data from continued experiments. Thus it was felt that results from an experiment such as the present one would provide a more valid estimate than any obtained previously.

If the present data are taken as more valid, then Stevens' (**4**) figure of 10-db. attenuation is almost certainly too low, since a value that low can be accepted on the basis of the present data only at a confidence level of less than .05.

Actually, there still remains some question as to whether the present experiment can be taken as one without bias. A considerable effort was made in the course of the instructions to *O* to be sure that he made a judgment with respect to half loudness rather than with respect to equal loudness, since the second tone was less loud than the first for all *O*s. Nevertheless, several of the *O*s used in this experiment later reported that they had originally judged with respect to equal loudness rather than half loudness, and their data were discarded. It is not known, of course, how many other *O*s may have done the same thing, but did not make any mention of it. This tendency will produce too many judgments of less than half as loud, with the result that the indifference point for half loudness will be too high (or too little attenuation from the standard). Thus even though this experiment gives a half loudness value

somewhat lower than Stevens suggests, it seems likely that the true value is even lower.

Summary

In this experiment, 135 different *Os* were required to make a single half-loudness judgment with the method of constants. They were required simply to state whether the second of a pair of tones was more or less than half as loud as the first, and they made this judgment on the first presentation of the tones without any preliminary practice in this situation. Different *Os* were given different intensities of the second (comparison) tone, but all had the same intensity for the first (standard) tone at 90 db. SPL.

A reasonable psychophysical function was obtained, and a least-squares solution of the normal ogive obtained. The half-loudness value thus obtained was at an attenuation of 12.4 db. from the standard intensity. It is suggested, however, that this attenuation value may be too low due to the fact that some *Os* judged with respect to equal loudness rather than with respect to half loudness.

References

1. Garner, W. R. Context effects and the validity of loudness scales. *J. exp. Psychol.*, 1954, **48**, 218–224.
2. Garner, W. R. A technique and a scale for loudness measurement. *J. acoust. Soc. Amer.*, 1954, **26**, 73–88.
3. Guilford, J. P. *Psychometric Methods.* (2nd Ed.) New York: McGraw-Hill, 1954.
4. Stevens, S. S. The measurement of loudness. *J. acoust. Soc. Amer.*, 1955, **27**, 815–829.
5. Stevens, S. S., & Poulton, E. C. The estimation of loudness by unpracticed observers. *J. exp. Psychol.*, 1956, **51**, 71–78.

(Received June 11, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

AGE CHANGES AND INFORMATION LOSS IN PERFORMANCE OF A PURSUIT TRACKING TASK INVOLVING INTERRUPTED PREVIEW[1]

STEPHEN GRIEW

Unit for Research on Employment for Older Workers, University of Bristol

Age changes in the performance of a pursuit tracking task in which no preview is given of the course to be followed have been studied by Welford (**10**). Using a "two-pointer" display, timing errors were found to be similar in both younger and older *S*s at high course speeds. At high speeds, breakdown in tracking was characterized by a reduction in the total displacement per course of *S*'s pointer, and this reduction was more pronounced among older *S*s. When reduction occurred, the total displacement of *S*'s pointer was inadequate to the task of maintaining perfect tracking, but the accompanying reduction in mean amplitude per response enabled *S*s to keep total misalignment and mean time lag at a minimum.

If preview of response requirements is given before matching takes place, it should be possible to reduce time lags by organizing responses in advance. Some evidence exists that this is so in simple serial tasks (**8**), and recent work by Welford[2] supports the view that it is so in pursuit tracking. By giving full preview, time lags in both younger and older *S*s were reduced substantially. Total displacement of *S*'s pointer also increases as preview increases, until an optimum duration of preview is given, above which further increases of preview are not accompanied by further improvements in performance (**11**). This optimum varies with age, from previews of duration .20 sec. in *S*s in their twenties to .50 sec. in *S*s in their sixties. At these previews it seems that amplitude error is at a minimum. This suggests that the combined effect of reaction time and refractory period, which appear to limit tracking performance when no preview is given (**6, 9. 10**), has decreased sufficiently to enable *S*s to maintain adequate timing without sacrificing amplitude accuracy.

It has been pointed out that the explanation of time lags in tracking in terms of the reaction time to two-choice stimuli is questionable, since the operator has to choose not only the direction of movement but also its extent (**7**). Recent work by Fitts (**2**) indicates that the time taken to organize movements depends largely upon their amplitude and tolerance limits, and this suggests further that the simple explanation of time lags in tracking must be modified. Crossman (**1**) has argued that amplitude errors in pursuit tracking reflect losses of information about the course to be followed. If a course is reproduced perfectly, albeit at a time lag, then loss of information cannot be said to have occurred.

On the basis of these findings and arguments, it is possible to recast the interpretation given by Welford and his colleagues (**10**) of their results. By increasing amplitude error, *S* may be said to be utilizing only a proportion of the information conveyed by the

[1] This study is part of a program supported by a grant from the Nuffield Foundation, London, the support of which is gratefully acknowledged. The author is indebted to G. C. Drew, E. R. F. W. Crossman, A. T. Welford, and K. F. H. Murrell, for advice and criticism.

[2] I am indebted to A. T. Welford, of the Psychological Laboratory, University of Cambridge, England, for this information.

course. In doing this, he reduces the choice level of separate stimuli, thereby decreasing his mean reaction time. In this way, by reducing his amplitude accuracy, he keeps his mean time lag at a minimum.

If preview is interrupted in a pursuit tracking task before matching takes place, separate responses will have to be delayed. If adequate time is given prior to interruption, amplitude errors in excess of those occurring in similar situations in which preview is uninterrupted may be said to be due to loss of information during storage occasioned by the need to delay responses. It seems necessary to distinguish between losses of information due to its storage and losses due to its nonutilization. This distinction appears to have implications for models of the operator which involve storage elements, and any observed age differences in losses of information due to storage would appear to be relevant to the much discussed issue of deterioration with age of "short-term" memory.

It appears that no examination has been made of age changes in the performance of a pursuit tracking task involving interrupted preview. Work on "prediction-motion" (**3, 4, 5**), in which the future activity of a course had to be predicted, does not provide examples of the type of experiment envisaged, and work on the limits of eye-hand span, although including tasks in which *S*s were required to respond at a time lag (**8**), was somewhat different from that proposed. These experiments were restricted, in any case, to younger *S*s.

In order to learn something of the age changes and the information loss due to storage occurring during performance of a pursuit tracking task involving interrupted preview, the following experiment was performed.

Method

The apparatus was that used by Welford (**11**) in studying uninterrupted preview. A complex harmonic course was generated by three epicyclic gears which actuated a side-to-side movement of a pen moving across a moving band of paper. The gears were driven by the same drive as the paper. The paper moved downward past a window placed perpendicularly to *S*s' line of sight. At the bottom of the window a pen was placed and this could be moved across the paper in tracking the course by means of a handwheel. A shutter was incorporated which could be raised or lowered to reveal more or less of the course before matching took place, and masks could be placed immediately above the pen to obscure the course at this point. With a mask in position, *S*s had to track the course at a point beyond that at which it had disappeared. Figure 1 shows the apparatus with a mask in position.

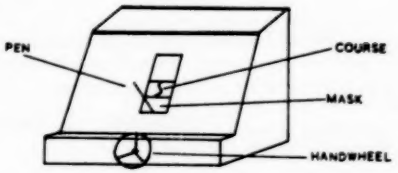

Fig. 1. The pursuit tracking apparatus.

Scoring was achieved by calculating mean timing errors and mean amplitude errors. Mean timing errors were obtained by converting into time values the mean vertical distances between course "peaks" and *S*s' track "peaks" by dividing these by course speed. Mean amplitude errors were computed by averaging the horizontal distances between "peaks."

Three levels of masking were combined with two course speeds to give six experimental conditions. Three masking levels (.00, 2.25, and 4.50 cm.) and two course speeds (4.50 and 9.00 cm./sec.) were studied. At all levels of masking, a preview above the mask of 4.50 cm. was given. This represented, at both speeds and for both groups of *S*s, at least optimum preview (**11**).

Six younger (aged 21–28 yr.) and six older (46–56 yr.) *S*s were randomly drawn from a large population of industrial workers of similar "background" and occupations, and were invited to take part in the experiment. None declined. The orders in which each *S* in each age group attempted the six experimental conditions were represented by the rows of a 6×6 Latin square. Before performance for experimental purposes, each *S* had one practice run at each of the experimental conditions. The length of paper passing *S*s' pens was the same at all conditions and involved no repetition of the course.

TABLE 1

MEAN TIMING ERRORS, AMPLITUDE ERRORS, AMPLITUDE ERRORS DUE TO STORAGE, AND RATE OF INFORMATION TRANSMISSION

Level of Masking (Cm.)	Course Speed (Cm./Sec.)	Timing Error (Sec.)		Amplitude Error (Cm.)		Resultant Ampl. Error (Cm.)		Info. Transmission Rate (Bits/Sec.)	
		Young *Ss*	Old *Ss*	Young *Ss*	Old *Ss*	Young *Ss*	Old *Ss*	Young *Ss*	Old *Ss*
.00	4.50	.06	.06	.20	.23			4.15	3.84
	9.00	−.07	−.05	.36†	.41†			5.81	5.31
2.25	4.50	.18	.17	.35‡	.56‡	.15	.33	2.97	1.97
	9.00	.07	.06	.48‡	.71‡	.12	.30	4.59	2.94
4.50	4.50	.37*	.30*	.41‡	.69‡	.21	.46	2.66	1.52
	9.00	.14	.14	.55‡	.80‡	.19	.39	4.00	2.44

Note.—Resultant amplitude error obtained by subtracting error at zero masking level.
* Difference between two mean timing errors is significant at $P = .05$. All other timing error differences between younger and older *Ss* are nonsignificant.
† Differences between mean amplitude errors of younger and older *Ss* are significant at $P = .05$.
‡ Differences between mean amplitude errors of younger and older *Ss* are significant at $P = .01$.

RESULTS AND DISCUSSION

Results, in terms of mean timing errors and mean amplitude errors for groups of *Ss*, are given in Table 1. Analyses of variance performed on timing error and amplitude error scores exhibited by separate groups of *Ss* demonstrated a significant influence of both Level of Masking and Course Speed in all cases. In order to determine the effect on amplitude error of storage, the mean amplitude error scores achieved when no masking was involved have been subtracted from the scores achieved, at the same course speeds, under masking conditions. The results of these subtractions are also given in Table 1.

Timing errors.—The small errors at the two no-masking conditions are similar to those found normally in pursuit tracking involving uninterrupted preview. The errors at masking conditions are all positive, indicating phase shifts in advance of the course, and follow much the same pattern in both younger and older *Ss*. The errors appear to increase as the duration of storage increases.

Amplitude errors.—Prima facie, these errors appear to increase as both extent of masking and course speed increase. This suggests that both the amount of information to be stored and the rate of storing and reproducing information are important determinants of information loss. This finding is similar to one reached on the basis of experiments by Poulton (**8**) referred to above. When the results on amplitude error due to storage are considered, however, it becomes evident that loss due to storage is determined primarily by the amount of information which has to be stored in delaying responses. The increase in error due to course speed is roughly the same, at both masking levels, as that which occurs with speed increase when no masking is involved, and is probably due, therefore, to the nonutilization of information in the first place. The information lost because of storage seems to be substantially higher in the case of the older *Ss* than in the case of the younger *Ss*.

In order to assess the extent of information loss due to storage it is useful to calculate the approximate mean rate of gain of information at each experimental condition. This may be done by freely adapting Crossman's method (1).

If A is the mean amplitude of course "swings," and ± Ae is the mean error exhibited in reproducing them, the probability of *S* achieving the mean accuracy exhibited may be said to be $\frac{2Ae.}{A}$ In this case, the mean amount of information used in achieving this degree of accuracy would be given by the expression $-\log_2 \frac{2Ae}{A}$ bits per "swing." If the number of "swings per course is S, the total amount of information used in achieving the mean accuracy exhibited will be equal to $-S \log_2 \frac{2Ae}{A}$ bits. If the duration of the course is D seconds, the rate of transmission of information will be equal to

$$\frac{-S \log_2 \frac{2Ae}{A}}{D} \text{ bits per second.}$$

The mean amplitude of course "swings" was 2.83 cm. The duration of the course at the lower speed was 64 sec. At the higher speed the course duration was 32 sec. In accordance with the method suggested above, the approximate rates of transmission of information at each condition have been calculated for the two groups of *S*s. These rates are given in Table 1. These data enable an estimate to be made of the loss of information per second due to storage.

The course speeds and their associated course durations varied very slightly during the experiment. The stated speeds and durations are modal, and the variation amounted to approximately ±2% of the stated values.

Summary

Information loss in pursuit tracking due to the nonutilization of information conveyed by the course to be followed is distinguished from information loss due to storage. An experiment is described in which *S*s are required to track a course which had disappeared from view, and the performances of younger and older *S*s are compared. Information loss due to storage appears to be determined mainly by the amount of information which has to be stored, and is severer in the case of older *S*s. Timing errors are similar in both younger and older *S*s and appear to be related mainly to the duration of storage. An attempt is made to express information loss due to storage in terms of a reduction in rate of information transmission.

References

1. Crossman, E. F. R. W. The measurement of perceptual load in manual operations. Unpublished Ph.D. thesis, Univer. of Birmingham, England, 1956.
2. Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *J. exp. Psychol.*, 1954, **47,** 381–391.
3. Gottsdanker, R. M. Prediction-motion with and without vision. *Amer. J. Psychol.*, 1952, **65,** 533–543.
4. Gottsdanker, R. M. The accuracy of prediction-motion. *J. exp. Psychol.*, 1952, **43,** 26–36.
5. Gottsdanker, R. M. A further study of prediction-motion. *Amer. J. Psychol.*, 1955, **68,** 432–437.
6. Hick, W. E. The discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1,** 36–51.
7. Hick, W. E., & Bates, J. A. V. *The human operator of control mechanisms.* London: Ministry of Supply, 1950.
8. Poulton, E. C. Eye-hand span in simple serial tasks. *J. exp. Psychol.*, 1954, **47,** 403–410.
9. Vince, M. A. The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, 1948, **38,** 149–157.
10. Welford, A. T. *Skill and age: an experimental approach.* London: Oxford Univer. Press, 1951.
11. Welford, A. T. Psychological aspects of aging. In W. Hobson (Ed.), *Modern trends in geriatrics.* London: Butterworth, 1956.

(Received June 19, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

EFFECTS OF PREVIOUSLY ASSOCIATED ANNOYING STIMULI (AUDITORY) ON VISUAL RECOGNITION THRESHOLDS [1]

JULIAN HOCHBERG AND VIRGINIA BROOKS

Cornell University

There is some evidence indicating that prior and simultaneous association with unpleasant stimuli may raise the visual recognition duration threshold of tachistoscopically presented words or shapes (**1, 3, 4, 6**). The purpose of the present experiment was to determine whether analogous effects could be obtained under nontachistoscopic presentation conditions. Two out of four simple geometric figures (a, b, c, d, adapted from Gottschaldt [**2**]) were separately paired with an annoying auditory stimulus (cf. **5**), and then all four figures were separately embedded in concealing test patterns (A, B, C . . ., P). The relative recognition brightness contrast thresholds of the concealed forms were measured by raising the illumination level for those figures, while keeping constant the remainder of the concealing test patterns.

METHOD

The apparatus consists of a visual system for presenting the concealed figures and the concealing patterns, and an auditory system for presenting the annoying stimulus.

The visual system consists of the following: (*a*) four polygonal training figures (**2**) in "white" on black (photostatic negatives), approximately 2 in. across.[2] (*b*) Sixteen test patterns, also white on black, four for each training figure, arranged so that the training figure is concealed by "good continuation," etc., among the lines of the test figure. The back of each test pattern is blacked over with India ink, except those lines which comprise the concealed figure. Thus, if illuminated from the front alone, the patterns only are visible; if illuminated only from behind, only the training figures transmit light and are visible; the relative brightnesses of the two sets of lines are thus varied by manipulating the relative intensities of front and rear illumination. (*c*) A stand holding the figures and patterns, illuminated by a front 50-w. lamp at 5 ft. (*d*) A 200-w. projector (TDC, A-2), 18 in. behind the stand. (*e*) An Eastman circular optical density wedge calibrated from minimum transmission at setting 100.0 to maximum transmission at setting 0.0. (*f*) A hand-operated fall shutter between *S* and the stand, permitting exposures of 5 sec. for the test patterns, and 10 sec. for training figures (G. E. timer, T-48).

The auditory system consists of a length of magnetic tape bearing selected bursts of chalk-on-blackboard screeches, somewhat over 10 sec. in overall duration; a tape recorder was kept running and warm throughout the training session; when certain training figures were illuminated for 10-sec. intervals, a simultaneous playback of the taped screeches was made to *S* through a set of earphones. Declarations of its unpleasantness or distractiveness were obtained from all *S*s.

Two groups of Cornell graduate and undergraduate students were used as *S*s, each assigned randomly either to Group A or Group B, 10 *S*s per group.

The experimental procedure consisted of a training series and a test series as follows:

In the training series, *S*s (wearing earphones) were told that they would be shown four figures which they were to remember and recognize later when presented in the same size and posi-

[1] The research reported here is part of a series of investigations conceived in the course of discussions with Gardner Murphy and his collaborators in the Perceptual Learning Project at the Menninger Foundation, and was supported in part by the United States Public Health Service Project M-715(C); the senior author is a consultant to this project. Richard Walters had suggested, and is employing, the Gottschaldt figures for a somewhat similar purpose.

[2] The training figures a, b, c, and d were Gottschaldt embedded figures III, II, IV, and V, respectively (**2**).

tion. These training figures were presented three times, in the following order: *a*, *b*, *c*, *d*, for 10-sec. exposures, two of the figures being "punished" by the simultaneous presentation of the taped screech. For Group A, Figures *a* and *c* were punished; for Group B, Figures *b* and *d*.

In the test series, earphones were removed, and *S*s were instructed that no further noise would be presented and that 16 pictures would be shown, each containing only one of the four training forms, which they were to report when recognized. The back-lighting of each pattern started at a setting of 90.0 and increased by steps of 5.0 units until the concealed training figure was correctly identified and located in the correct quadrant of the test pattern.

Results and Discussion

For each *S*, the sum of wedge threshold settings for recognition of "punished" test forms was subtracted from the corresponding score for "nonpunished" forms. Seventeen of the 20 differences were positive, i.e., 85% of the *S*s displayed higher recognition thresholds for "punished" than "nonpunished" test forms. The differences ranged from −16.9 to 30.6, with an *SD* of 11.8, a mean of 6.95 and *t* of 2.6 ($P < .02$).

Summary

Visual recognition brightness contrast thresholds can be raised by pairing an annoying auditory stimulus with forms which are later embedded in more complex patterns, under nontachistoscopic exposure conditions.

References

1. Ayllon, T., & Sommer R. Autism, emphasis, and figure-ground perception. *J. Psychol.*, 1956, **11**, 163–176.
2. Gottschaldt, K. Uber den Einfluss der Erfahrung und die Wahrnehmung von Figuren. I. *Psychol. Forsch.*, 1926, **8**, 261–317.
3. Hochberg, J. E., Haber, S. E., & Ryan, T. A. "Perceptual defense" as an interference phenomenon. *Percept. mot. Skills*, 1955, **5**, 15–17.
4. McCleary, R. A., & Lazarus, R. S. Autonomic discrimination without awareness: an interim report. *J. Pers.*, 1949, **18**, 170–179.
5. Simpson, J. E. The influence of auditory stimulation on aniseikonic perception: a preliminary study. *J. Psychol.*, 1956, **41**, 235–241.
6. Smith, D. E. P., & Hochberg, J. E. The effect of "punishment" (electric shock) on figure-ground perception. *J. Psychol.*, 1954, **38**, 83–87.

(Received June 21, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

POSITION PREFERENCE AND DISCRIMINATION LEARNING[1]

MARVIN H. GOER

Michigan State University

This study is concerned with the effect that strengthening a position preference has on subsequent brightness-discrimination training. According to Spence's (4) analysis of simultaneous brightness-discrimination learning, the slow emergence and rapid approach to criterion of the to-be-learned response is due to the presence of a competing position habit. Thus, according to Spence, it is necessary to assume that the position habit is being extinguished while the brightness habit is being strengthened. Spence accounts for the rather slow initial learning of the brightness discrimination problem by assuming that competing position habits are strong in comparison to the brightness habits. Most studies have attempted to control the position variable rather than manipulate it.

Even though one accepts Spence's analysis with respect to competing position responses, it is not necessary to make the assumption that the position responses are initially stronger than the responses to brightness cues. Much of the strengthening of the position response may well occur during the initial stages of typical brightness-discrimination training. If an interference theory of extinction is accepted, then the neutralization of the position response requires the establishment of a competing (avoidant) response to the position cues. Furthermore, if the framework of elicitation theory (1) is accepted, a habit must be of some minimal magnitude before nonreward can elicit the competing avoidant responses. This means that the neutralization of the position habit through nonreward will be facilitated by first strengthening it to some degree.

The literature contains several instances in which nonreward is more effective when the response tendency is strong. This includes the observation by Spence (5) that nonreward was ineffective in discrimination learning with chimpanzees unless the response has previously been rewarded. A study by Denny and Dunham (2) showed that nonreward of an incorrect spatial response in a T maze had little effect unless there was a strong initial position preference for the incorrect response.

On the basis of the above considerations, the following hypothesis is formulated: moderate strengthening of a position preference prior to simultaneous brightness-discrimination training will facilitate brightness-discrimination learning, while strengthening the position preference to a high level will impede such discrimination learning.

METHOD

Subjects.—The *S*s were 46 naive hooded rats, 25 males and 21 females, 90–120 days old drawn from the colony of the psychology department at Michigan State University. The *S*s were assigned to their respective groups in a random fashion.

Apparatus.—The apparatus was a modified

[1] This paper constitutes a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Michigan State University. The writer wishes to express his sincere appreciation to M. Ray Denny for his direction and encouragement.

Grice-type discrimination box. The box was constructed of .5-in. plywood throughout and covered with hardware cloth. The starting chamber was a trapezoidal shaped box 8 in. wide at the base joined at its narrow end to an alley 2 in. wide and 4 in. long. The alley led into a trapezoidal reaction chamber and this chamber had an 18-in.-wide base opening into the two goal boxes presented on any one trial. The starting chamber, alley, and reaction chamber were 16 in. long from the rear wall of the starting chamber to the entrance of the goal boxes. Two guillotine doors 4 in. wide and 9 in. high could be lowered between the reaction chamber and the goal boxes. The goal boxes were three adjacent boxes of equal dimensions; any two adjacent boxes opened into the reaction chamber. The dimensions of each goal box were 4 in. wide and 8 in. deep. The over-all length of the apparatus from the rear wall of the starting chamber to the far wall of the goal box was 24 in. All the walls of the apparatus were 8 in. high.

The starting chamber, alley, reaction chamber, and the guillotine doors were unpainted. The goal boxes were painted as follows: The interior of the two outside boxes were painted flat white, and the interior of the center box was painted flat black. At the rear of each goal box was a baffle—a $5 \times 2 \times \frac{1}{2}$-in. piece of plywood attached to the left wall of the box 2 in. from the end wall and flush with the floor of the box. When a food pellet was placed behind this baffle it could not be seen by *S* from any point outside the box. A 200-w. bulb hung 5 ft. above the guillotine doors.

Habituation.—All *S*s were placed on a 24-hr. feeding schedule for at least three days prior to the start of the preliminary training. They were fed in their home cage and received a mixture of Purina dog chow (their normal diet) and Zinns food pellets (which were used as reward objects). An amount of food equal to 8 gm. per *S* was introduced into the home cage during this procedure at the same time each day.

In preliminary training, the *S*s were placed in an unpainted alley, the floor of which was littered with Zinn's food pellets. The alley was constructed of two end boxes, 4 in. wide, 6 in. high, and 8 in. long placed with their open ends together so that they composed an alley 16 in. long. The *S*s were placed in this alley for approximately 1-hr. prior to their normal feeding time. While in the alley they were handled individually by *E* and allowed free feeding of the food pellets. Following this they were replaced in the home cage which contained their daily ration. This procedure was followed for four days after which *S*s were placed in individual feeding cages which contained their daily ration for 1 hr. prior to being returned to their home cage in which only water was available. Position preference was determined on the day following habituation.

Determination of position preference.—The two unpainted end boxes which had served as the alley during habituation were placed in the goal box position of the apparatus with several pellets of food in each of them. Each *S* was introduced into the starting chamber and allowed to enter either the right or left goal box. This response was recorded and was designated as the position preference.

Running procedures.—As soon as *S* placed both of its hind feet into the goal box, the guillotine door was lowered. This was the criterion of a response and defined the trial. No retracing was permitted. If, after being placed in a starting chamber, *S* failed to respond within 6 min., it was removed from the apparatus and discarded. Fifteen such *S*s were discarded, most of them before Trial 5. Starting with position preference training and continuing on through discrimination learning, all *S*s received 3 trials on Day 1, 7 trials on Day 2, and 10 trials per day thereafter. On the first trial that *S* was rewarded, it was allowed to stay in the goal box for 6 min. or until it ate the food pellet. On all subsequent trials, each *S* remained in either end box for 30 sec. The *S*s were run in rotation and the last *S* to reach criterion was delayed at least 60 sec. between trials. A reward pellet weighed approx. .05 gm. The *S*s were run under 22 hr. food deprivation and were placed in individual feeding cages and allowed 1 hr. free feeding following the day's running before being replaced in the home cage.

Position-preference training.—On the day following position-preference determination, *S*s were divided into two control (C) and three experimental (E) groups. The black and white goal boxes were now placed in position and the white goal box was alternated in the following order: LRLRRLRLLR so that each color appeared equally often on the preferred side.

Group C-O ($N = 10$) received no experimental strengthening of the position preference. On the first day following position-preference determination, the positive cue for this group was either white or black. This condition prevailed until each *S* achieved the criteria. Group E-12 ($N = 10$) was given 12 trials with the preferred side rewarded. Food reward was always in the right or left goal box. The number of rewards received depended upon the number of responses to the preferred side. Following this training *S*s were switched to brightness-discrimination learning (food reward was always in the black or white goal box).

TABLE 1

t RATIOS FOR COMPARISONS BETWEEN GROUPS

Groups	C-24	E-12	E-24	E-48
C-O	.047	3.21*	3.19*	3.93*
C-24		3.14*	3.08*	4.13*
E-12			.534	5.89*
E-24				5.90*

* $P < .01$.

Group E-24 ($N = 10$) received 24 trials with the preferred side rewarded and then was switched to brightness-discrimination learning. Group C-24 ($N = 10$) was matched with Group E-24 so that every *S* in Group C-24 was rewarded on the same trial as its match-mate of Group E-24 irrespective of position or brightness. This was accomplished by placing reward objects in both goal boxes on the to-be-rewarded trial. If S_1 in Group E-24 received reinforcement on Trials 1, 5, 6, 9, then S_1 in Group C-24 received reinforcement on Trials 1, 5, 6, 9. In other words, Group C-24 received the same number of trials, number of rewards, and pattern of rewards, prior to brightness-discrimination learning that Group E-24 did but with no consistent strengthening of the position preference. Group E-48 ($N = 6$) received 48 trials with the preferred side rewarded and then was switched to brightness-discrimination learning.

Discrimination training.—On the day of the last block of position-preference training trials, all *S*s were shifted to black-white discrimination training. For C-O this occurred on the first day following position-preference determination. This meant that E-12 received eight, E-24 received six, C-24 received six, and E-48 received two black-white discrimination trials on this day. Subsequently each *S* was run 10 trials per day. The criterion of learning was 9 out of 10 correct responses in any one day. Black was positive for one-half the *S*s in each group and white was positive for the other half.

RESULTS AND DISCUSSION

The mean number of trials to criterion including criterion trials was 56 ($SD = 10.19$) for Group C-O, 55 ($SD = 9.20$) for Group C-24, 38 ($SD = 13.41$) for Group E-12, 41 ($SD = 10.24$) for Group E-24 and 85.3 ($SD = 14.90$) for Group E-48.

The requirement of homogeneity of variance was satisfied and an *F* test indicated that there was a significant difference between groups ($F = 17.46$, $df = 4$ and 41, $P < .01$). Table 1 presents the *t*'s for individual comparisons between groups. Table 1 shows that Groups C-O and C-24 were not significantly different from one another and that Groups E-12 and E-24 were not significantly different from one another. However, Groups E-12 and E-24 performed significantly better than both control groups. Group E-48, on the other hand, performed significantly poorer than all other groups. The general pattern of these results conform to the present hypothesis. The fact that Group C-24 did not learn any faster than Group C-O is of particular importance. This indicates that experience with the differential cues and the apparatus does not account for the superiority of the E-12 and E-24 groups.

Since *S*s were rewarded for spatial responses in the position-preference training, it is necessary to find out whether there was any differential incidental strengthening of the to-be-positive brightness cue. Specifically, we find that Group C-24 was rewarded 6.8 times in the to-be-positive goal box and Group E-24 7.5 times. This difference of .7 rewards is not significant ($t = .65$). Furthermore, the correlation between number of rewards on the to-be-positive cue during position-preference training and trials to criterion for all *S*s was negligible (eta = .05, $r = .04$).

As far as the writer can determine, the results obtained in this experiment cannot be accounted for by any existing theory of discrimination learning but can be interpreted within the broader framework of elicitation theory. Specifically the postulate of "secondary elicitation" states that the removal of a

reward object in an *established* behavior sequence will elicit characteristic avoidant responses which become conditioned to the cues associated with nonreward. In further support of this analysis Shoemaker (**3**) in a recent study has found that the effects of nonreward to a brightness cue are enhanced by first giving 12 rewarded trials to this cue.

Summary

Rats ($N = 46$) were trained in a Grice-type discrimination box. Three experimental groups (E) were given differential amounts of position-preference strengthening (12, 24, and 48 trials with the preferred side positive). Two control groups (C) were given 0 and 24 inconsistently rewarded prediscrimination trials. All groups were then shifted to a black-white discrimination problem and trained to a criterion of 9 out of 10 correct on a day's run of 10 trials. Groups E-12 and E-24 performed significantly better than did each of the other groups, and Group E-48 performed significantly poorer than each of the other groups. There was no difference between Groups E-12 and E-24 or C-0 and C-24. These results were consistent with the present hypothesis and were interpreted in terms of elicitation theory.

References

1. Denny, M. R., & Adelman, H. M. Elicitation theory: I. An analysis of two typical learning situations. *Psychol. Rev.*, 1955, **62,** 290–296.
2. Denny, M. R., & Dunham, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in a simple T-maze. *J. exp. Psychol.*, 1951, **41,** 382–389.
3. Shoemaker, H. A. The relative efficiency of rewarded and nonrewarded training in a black-white discrimination problem. Unpublished Ph.D. thesis, Univer. of Colorado, 1953.
4. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43,** 427–499.
5. Spence, K. W. Analysis of the formation of visual discrimination habits in the chimpanzee. *J. comp. Psychol.*, 1937, **23,** 77–100.

(Received June 24, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

THE EFFECTS OF DEGREE OF LEARNING AND PROBLEM DIFFICULTY ON PERSEVERATION[1]

ROBERT THOMPSON

Southeast Louisiana Hospital

It has been repeatedly demonstrated that the interpolation of such treatments as ECS or anoxia selectively disrupts memories of newly acquired habits (**1, 4, 8, 10**). These data provide strong evidence for the Müller and Pilzecker (**2**) perseveration theory which states that the complete establishment of the memory trace occurs at some time following training through the persistence of certain neural events. The period through which these hypothetical events act (perseveration time) may be approximated by determining the maximal interval after training at which the interpolation of an ECS still produces a significant deficit in memory (**5**).

It might be expected that perseveration time (PT) will be a function of a number of variables. The results obtained by Thompson and Pennington (**9**), for example, suggest that PT decreases with increasing degrees of distributed practice in the original learning of the habit. In the present study, the relative effects upon PT of both the level of learning and the difficulty of the problem were investigated. Due to the character of these variables, two separate experiments were performed. Experiment I examines the effects of four levels of learning, while the effects of three problems differing in difficulty are reported in Exp. II.

[1] This study was conducted at Louisiana State University and supported, in part, by a grant-in-aid from the Graduate Research Council.

EXPERIMENT I

Method

Subjects.—The *Ss* were 160 male albino rats of the Harlan strain. At the beginning of experimental training, *Ss* ranged in age from 75 to 160 days. All *Ss* were kept in individual cages containing a constant supply of food and water. Handling was practiced three days prior to the onset of preliminary training.

Apparatus.—A two-choice discrimination box was employed which utilized the motive of avoidance of electric shock. A detailed description of the apparatus has been presented elsewhere (**7, 8**).

The cerebral shock unit was the same as that used in a previous study (**8**). A current strength of approximately 50 ma. lasting for .5 sec. was administered through alligator clips mounted on *Ss'* ears.

Procedure.—Preliminary training which was carried out over a period of five days was the same as that previously described (**10**).

The test problem involved the discrimination between horizontal and vertical stripes, the vertically striped card constituting the positive stimulus for all *Ss*. The thickness of the alternating black and white striations was ¼ in. A response to the positive card admitted *S* to the goal box, whereas a response to the locked negative card was punished. Punishment consisted of receipt of an electric shock from the grid section located below the negative stimulus card. The correction technique was used.

Four main groups of rats, each containing 40 *Ss*, were given 10, 20, 30, and 40 trials, respectively, on the test problem. The intertrial interval was 45 sec. The positive card was switched from the right to the left window in a prearranged random order. The specific training procedure was as follows. The *S* was placed into the start box and the door raised. Failure to leave the start box within 5 sec. was followed by a brief shock from the grid floor. No further shock was given unless *S* made an error. An error was defined as an approach response to the negative card which brought *Ss'* forefeet in contact with the charged grid section. Throughout this study, only initial errors were recorded.

Each main group was divided at random into

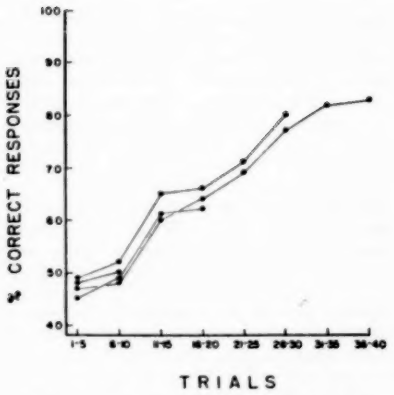

FIG. 1. The performance curves for the four main groups.

four subgroups of 10 *Ss* each. One subgroup from each main group received a convulsive shock 30 sec. after the completion of the last training trial. The second and third subgroups from each main group were given an ECS 3 min. and 30 min., respectively, after completing the last training trial. The remaining subgroup from each main group constituted the controls to which no ECS was given.

Twenty-four hours later, all *Ss* were given 30 trials on the same discrimination problem (45-sec. intertrial interval). The training procedure was identical to that involved in initial training. The number of errors made during this postshock period was utilized as the index of retention.

Results

Figure 1 presents the separate performance curves derived from the four main groups on the initial training trials. In terms of performance on the last 10 trials, the level of learning attained by the 10-trial group was at chance, while the 20-, 30- and 40-trial groups were approximately at the 60%, 75%, and 80% levels, respectively.

TABLE 1

MEAN ERRORS ON 30 POST-SHOCK TRIALS FOR ALL SUBGROUPS

Shock Cond.	Trials			
	10	20	30	40
30 sec.	11.0	9.5	7.4	4.5
3 min.	9.3	8.3	6.6	2.9
30 min.	7.6	7.7	5.5	3.3
Control	8.0	5.6	4.3	3.2

The mean error scores for all subgroups on the postshock training trials are summarized in Table 1. The convulsed *Ss* tended to be inferior to the controls for all levels of learning, this trend being particularly appreciable in the 20-trial group. However, the information in Table 2 shows that ECS produced significant retroactive effects when administered either after 10, 20, or 30 trials, but failed to cause any significant interference when administered after 40 trials.

For the purpose of estimating the duration of perseveration for each level of learning, the differences between the experimental and corres-

TABLE 2

ANALYSES OF VARIANCE OF POSTSHOCK ERROR SCORES FOR EACH LEVEL OF LEARNING

Source of Variation	*df*	10 trials		20 trials		30 trials		40 trials	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Between groups	3	23.53	3.26*	26.60	4.83**	16.25	3.37*	5.15	1.37
Within groups	36	7.21		5.50		4.82		3.76	
Total	39								

* $P = .05$.
** $P = .01$.

ponding control subgroups were subjected to a t test. With respect to the 10-trial group, the *Ss* given an ECS 30 sec. after training made more errors than the controls ($P = .02$). The two other shock conditions, however, failed to cause any significant memory deficits. For the 20-trial group, all subgroups receiving an ECS were significantly inferior to the controls on the postshock trials ($P < .05$). Only the 30-sec. and 3-min. shock conditions produced significant deficits for the 30-trial group ($P = .01$ and .05, respectively). Concerning the 40-trial group, none of the differences between the experimentals and the controls approached statistical significance.

To summarize, these data suggest that, during the early phase of learning, PT increases with increasing amounts of practice. This was shown by the fact that *Ss* given 10 trials revealed a PT of about 30 sec., whereas *Ss* given 20 trials revealed a PT of at least 30 min. With additional training, however, it appears that there is a decrease in PT. Thus, 30 trials yielded a PT of about 3 min., while a PT of less than 30 sec. was produced by 40 trials.

Experiment II

Method

Subjects and apparatus.—The *Ss* were 90 experimentally naive albino rats of the Harlan strain, of which 60 were females. Their ages at the beginning of preliminary training ranged from 50 to 110 days.

The discrimination apparatus and convulsive shock unit were the same as those used in Exp. I.

Procedure.—On Days 1 and 2, each *S* was allowed to explore the goal box for periods of 15 min. Cards placed against the windows prevented *S* from entering the choice chamber.

On Day 3, *S* was trained to run from the start box through the choice chamber and into the goal box in order to avoid shock, while medium gray cards were placed against the windows through which each *S* was trained to run on Day 4.

On Day 5, *Ss* were randomly divided into three main groups of 30 *Ss* each. One group was trained on a position habit (Problem P), while the other two groups received training on one of two visual discrimination habits (Problems D_1 and D_2). With respect to Problem P, *Ss* were trained positively to the nonpreferred side. Black cards were placed against both windows for half of the *Ss*, while white cards were used for the remaining half. The card on the preferred side (negative) was locked throughout training. Problem D_1 was a brightness discrimination involving a white card (positive) and a black card (negative). Problem D_2 was the same as Problem D_1, save for the negative card which consisted of alternating black and white vertical stripes, the thickness of the striations being $\frac{1}{2}$ in. The positive card was randomly switched from left to right during training for both visual discrimination problems.

For all problems, a response to the positive stimulus (card or position) admitted *S* to the goal box, whereas a response to the locked negative stimulus (card or position) was punished. Punishment involved receipt of an electric shock from the grid section located below the negative stimulus. The specific training procedure was the same as in Exp. I. Using an intertrial interval of 45 sec., training on Day 5 was terminated when *S* reached the criterion of five consecutive errorless trials.

Each main group was divided into three equal subgroups of 10 *Ss*. Two subgroups from each main group were given an ECS 20 sec. and 15 min., respectively, following the completion of the last criterion trial. The remaining subgroup from each main group constituted the controls to which no ECS was given.

On Day 6, all *Ss* were given additional trials (45-sec. intertrial interval) on the same problem that was mastered on the previous day. The training procedure was identical to that involved in original learning. The *Ss* having Problem P received 10 trials, while *Ss* having either Problem D_1 or Problem D_2 received 20 trials. The number of errors committed on Day 6 was used as the index of retention.

Results

Table 3 presents the original learning scores for all subgroups. It is apparent that the three problems were of unequal difficulty. Use of the t test showed that Problem P was significantly easier than Problem D_1 ($P < .01$ for trials to criterion and errors to criterion). Similarly, Problem D_2 was more difficult than

Problem D_1 ($P < .01$). To determine whether there were any significant differences in initial learning speed among the subgroups trained on the same problem, an analysis of variance was applied to both criterion measures. All F's were far from statistical significance.

The error scores for all subgroups on the relearning trials can be seen in Table 4. With respect to the 20-sec. shock condition, t's for the differences between the experimental *Ss* and the corresponding controls revealed that ECS resulted in significant retroactive inhibition on Problems D_1 and D_2, but failed to cause any interference on Problem P. Furthermore, it is to be noted that while the 15-min. shock condition produced no apparent memory deficit on Problem D_1, it did cause a marginally significant deficit on Problem D_2, ($P = .10$). Thus, PT seems to increase with problem difficulty, varying from less than 20 sec. for Problem P to about 15 min. for Problem D_2.

TABLE 3

MEAN TRIALS AND MEAN ERRORS TO CRITERION FOR ALL SUBGROUPS

Shock Cond.	Problem P		Problem D_1		Problem D_2	
	Trials	Errors	Trials	Errors	Trials	Errors
20 sec.	2.8	2.4	15.5	7.6	40.2	18.8
15 min.	2.8	2.0	16.8	7.3	34.4	15.2
Control	2.4	2.0	16.4	8.1	33.2	15.6

DISCUSSION

The essential findings of these experiments are that the level of learning as well as the difficulty of the problem varies the degree to which ECS disrupts memory. This has been shown by consideration of perseveration time which represents a measure of retroactive inhibition derived from the classical Müller and Pilzecker theory. Other factors which have been found to influence the duration of perseveration are the length of the intertrial interval (**9**) and the age of *S* (**6**).

As previously discussed (**9, 10**), the effects of these variables on PT may perhaps best be explained in terms of the differential accrual of perseverative activity from trial to trial. It is apparent, however, that such an explanation is only post hoc, and that adequate predictions based upon the theory must await formal statements concerning the organic nature of perseveration. Conceivably, insight into the perseverative process may come with the investigation of PT as affected by physiological variables. That such variables exist has already been shown by Pennington (**3**), who recently found that PT is longer in brain-damaged rats than in normals.

SUMMARY

This study aimed to determine the relative effects of the level of learning and the difficulty of the problem on perseveration time in rats.

TABLE 4

MEAN ERRORS ON RETENTION TRIALS FOR ALL SUBGROUPS

Shock Cond.	Problem P			Problem D_1			Problem D_2		
	M	Diff.	*t*	*M*	Diff.	*t*	*M*	Diff.	*t*
20 sec.	1.3	.8	1.54	3.4	1.7	2.68*	7.7	3.5	3.30**
15 min.	.5			1.9	.2	<1.00	6.6	2.4	1.91
Control	.5			1.7			4.2		

* $P = .02$.
** $P = .01$.

The results of Exp. I suggest that PT increases early in practice and is then followed by a decrease with further practice. In Exp. II, PT was found to increase with the difficulty of the problem.

References

1. Duncan, C. P. The retroactive effect of electroshock on learning. *J. comp. physiol. Psychol.*, 1949, **42,** 32–44.
2. Müller, G. E., & Pilzecker, A. Experimentelle Beiträge zur Lehre vom Gedächtniss. *Z. Psychol.*, 1900, Ergbd. 1, 1–288.
3. Pennington, D. F. The effect of brain damage on retention following ECS. Unpublished doctoral dissertation, Louisiana State Univer., 1957.
4. Ransmeier, R. E., & Gerard, R. W. Effects of temperature, convulsion and metabolic factors on rodent memory and EEG. *Amer. J. Physiol.*, 1954, **179,** 663–664. (Abstract)
5. Thompson, R. The comparative effects of ECS and anoxia on memory. *J. comp. physiol. Psychol.*, 1957, **50,** 397–400.
6. Thompson, R. The effect of ECS on retention in young and adult rats. *J. comp. physiol. Psychol.*, 1957, **50,** 644–646.
7. Thompson, R., & Bryant, J. H. Memory as affected by activity of the relevant receptor. *Psychol. Rep.*, 1955, **1,** 393–400.
8. Thompson, R., & Dean, W. A further study on the retroactive effect of ECS. *J. comp. physiol. Psychol.*, 1955, **48,** 488–491.
9. Thompson, R., & Pennington, D. F. Memory decrement produced by ECS as a function of distribution in original learning. *J. comp. physiol. Psychol.*, 1957, **50,** 401–404.
10. Thompson, R., & Pryer, R. S. The effect of anoxia on the retention of a discrimination habit. *J. comp. physiol. Psychol.*, 1956, **49,** 297–300.

(Received July 2, 1957)

Journal of Experimental Psychology
Vol. 55, No. 5, 1958

CONDITIONING OF THE HUMAN PUPILLARY DILATION RESPONSE AS A FUNCTION OF THE CS-UCS INTERVAL [1]

ARNOLD A. GERALL [2] AND JOHN K. WOODWARD

University of Rochester

This study was an attempt to determine the relative effect of four different CS-UCS intervals upon the classical conditioning of the pupillary dilation response. This information was sought primarily for its import for the design of experiments using the dilation response and, secondarily, for its possible bearing upon hypotheses concerning the nature of the modification produced by the classical conditioning method.

No studies could be found in the literature in which the CS-UCS interval had been varied systematically when conditioning the dilation response. A study has been performed by Hilgard, Dutton, and Helmick (**6**), in which four CS-UCS intervals (0, 3, 6, and 9 sec.) were used in an attempt to condition pupillary constriction. They reported that only one of their 43 *Ss* showed any CR's. It might be concluded that these intervals were equally ineffective and other values should be presented in future investigations. Their negative findings, however, might not have been due to these particular intervals but to the absence of other variables or conditions necessary for the modification process. In addition, the parameters of the variables characteristic of the constriction response might not be the same as those relevant to dilation activity. For these reasons, the results of the cited study by Hilgard, Dutton, and Helmick do not facilitate a selection of CS-UCS values for the conditioning of the dilation response.

It was on the basis of the experiments of White and Schlosberg (**10**) and of Moeller (**9**) that a choice was made of the intervals to be investigated. In these studies, CS-UCS separations ranging from 0 to 4 sec. were utilized and the level of GSR conditioning was recorded. The optimal interval was found to be approximately .5 sec. Practically no conditioning was found when the intervals were greater than 2.0 sec. Since the GSR and dilation responses could be evoked by similar stimuli and because both are innervated by the autonomic nervous system, values from the above studies were selected for this experiment.

In discussing the implications of their results for various learning theories, White and Schlosberg suggest that an S-R position would require the optimal interval to be related to the recorded latency of the UCR. An S-S view would maintain, on the other hand, that the optimal interval is dependent on the temporal relationships between the CS and UCS and independent of any response latencies. The mean latency of the GSR is about 2 sec. and the mean latency of a skeletal response, such as the eyeblink, is in the order of .04 sec. (**10, 11**). Because the duration of the optimal CS-UCS interval for the conditioning of the GSR and the eyeblink are approximately the same

[1] This research was supported by the National Science Foundation under Grant NSF-61791.

[2] Now at the Department of Anatomy, University of Kansas.

(7, 8, 9, 10), White and Schlosberg state that the results from these experiments are encompassed best by an S-S theory.

The latency of the pupillary dilation response to an illumination change under stable conditions has been found to be about .180 sec., and following pupillary activity, about .385 sec (12). The latency of the response to shock has been reported to be approximately .29 sec. (4). The latency of the pupil to combination of light decrease and shock which was the UCS used in the present study has not been systematically determined and is probably related to the level of shock and to intra-*S* conditions. But judging from the data noted above and some observations made in an earlier study (2), it is certain that the latency of the dilation response to a complex stimulus of shock and light offset is shorter than the latency of the GSR and longer than that of the eyeblink. According to an S-S theory, the difference in latency of the pupil should not influence the conditioning process, and the optimal CS-UCS interval of .5 sec. should also be obtained in this study.

Method

Subjects.—Thirteen men and 12 women undergraduates from the University of Rochester served as *S*s. They were randomly assigned to each of the five experimental groups.

Apparatus.—The apparatus used in this study has been described in detail elsewhere (2). In brief, infrared motion picture photography was used to record the diameter of the pupil. The camera speed was 10 frames per second. Both eyes were illuminated by the infrared and UCS light sources, but only the right eye was photographed.

The CS consisted of the cessation of thermal noise and the onset of a 1000 cy./sec. tone, approximately 50 db above threshold, generated by a commercial oscillator. Both stimuli were delivered through earphones in an aviator's helmet which *S* wore. Except during rest periods and when the CS was presented, thermal noise, approximately 70 db above threshold, was transmitted through the earphones. This noise was sufficiently loud to mask the sound of switches and other sounds.

The UCS was a combination of light offset and electric shock. The lamp whose offset served as part of the UCS was approximately 30 ft.-candles, and when it was turned off during the UCS period, a fraction of a foot-candle was available to *S*. The shock was a 60-cycle voltage conducted through silver electrodes located on the *S*'s right forearm. The voltage level, which ranged from 10 to 40 v., was established for each *S* by increasing the voltage in 2-v. steps until the shock was reported to be uncomfortable and something that he would avoid.

Procedure.—A delayed conditioning procedure with the CS and UCS terminating together was used. The intervals between the onsets of the CS and UCS presented to each group were as follows: Group I—.125 sec.; Group II—.50 sec.; Group III—1.50 sec.; Group IV—2.50 sec.; and Group V was a sensitization control. The UCS duration was always 1.50 sec. and the camera was on for the total trial period. Each group consisted of 5 *S*s.

The general experimental procedure was the same for all *S*s. There were two experimental sessions, usually on consecutive days but sometimes separated by one day. On Day 1, *S* was given instructions and the equipment was adjusted for photography, but no photographs were taken. The CS and UCS were never paired on this day but were presented separately. During Day 2, the experimental period consisted of four adaptation trials, 20 conditioning trials with four test trials interspersed, and 8 extinction trials. The *S*s in the conditioning groups had the following order of experimental stimuli: shock level adjusted, A (adaptation trial), A, A, A, C (conditioning trial), T (test trial), C, C, C, C, T, C, C, C, C, C, T, C, C, C, C, C, C, T, C, C, C, E (extinction trial), E, E, E, E, E, E, E. The sensitization control group was presented with the same order but on conditioning trials the UCS was presented alone. The intertrial interval was varied randomly, but arranged to average out to 1 min. The pupil was photographed on all adaptation and test trials, but only on the first, third, fifth, and eighth extinction trials. The diameter of the pupil was measured on a microfilm reader that magnified the image on the 16-mm. film 27 times. This value was converted back to actual size in millimeters by dividing by an appropriate conversion factor.

Results

The results are based upon the diameter of the pupil measured on successive frames exposed every tenth second during the CS interval. The average diameter of the pupil at each tenth second was computed for each group for their recorded adaptation, test, and extinction trials. Four trials were photographed during each of the

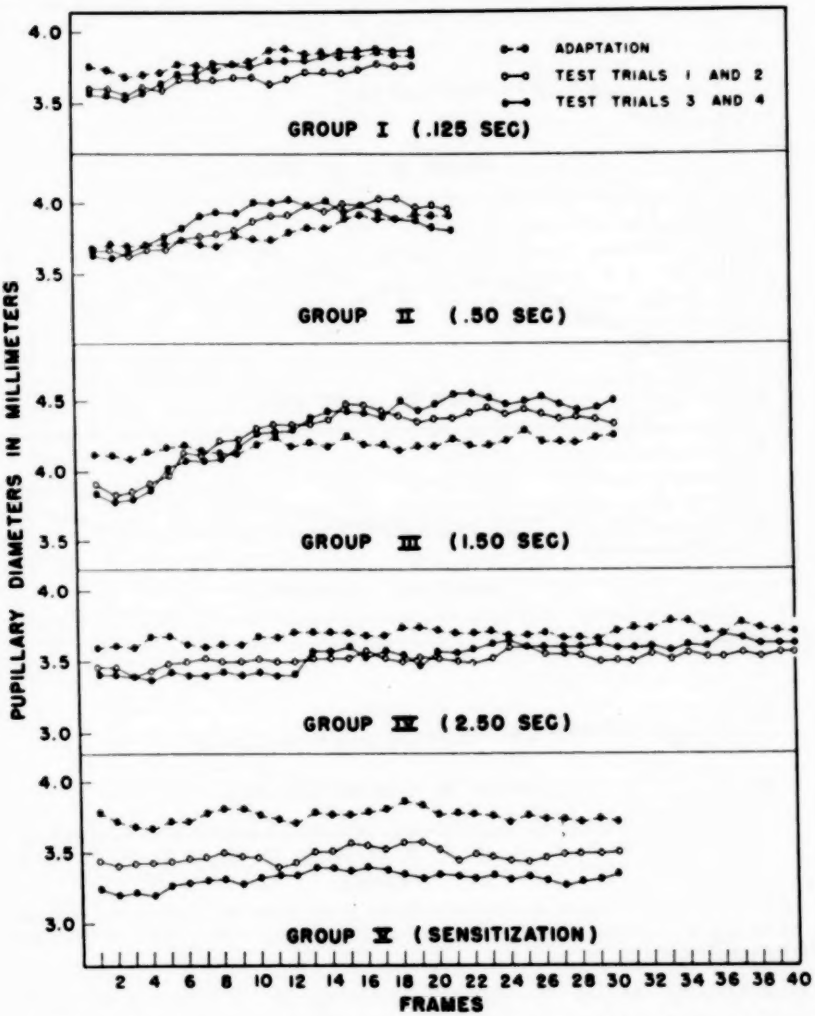

FIG. 1. The mean diameter in millimeters of the pupil during adaptation and test trials for each of five groups. The first frame coincided with the onset of the CS and 1/10 sec. elapsed between each succeeding frame. The order of presentation of the test trials is provided in the description of procedure.

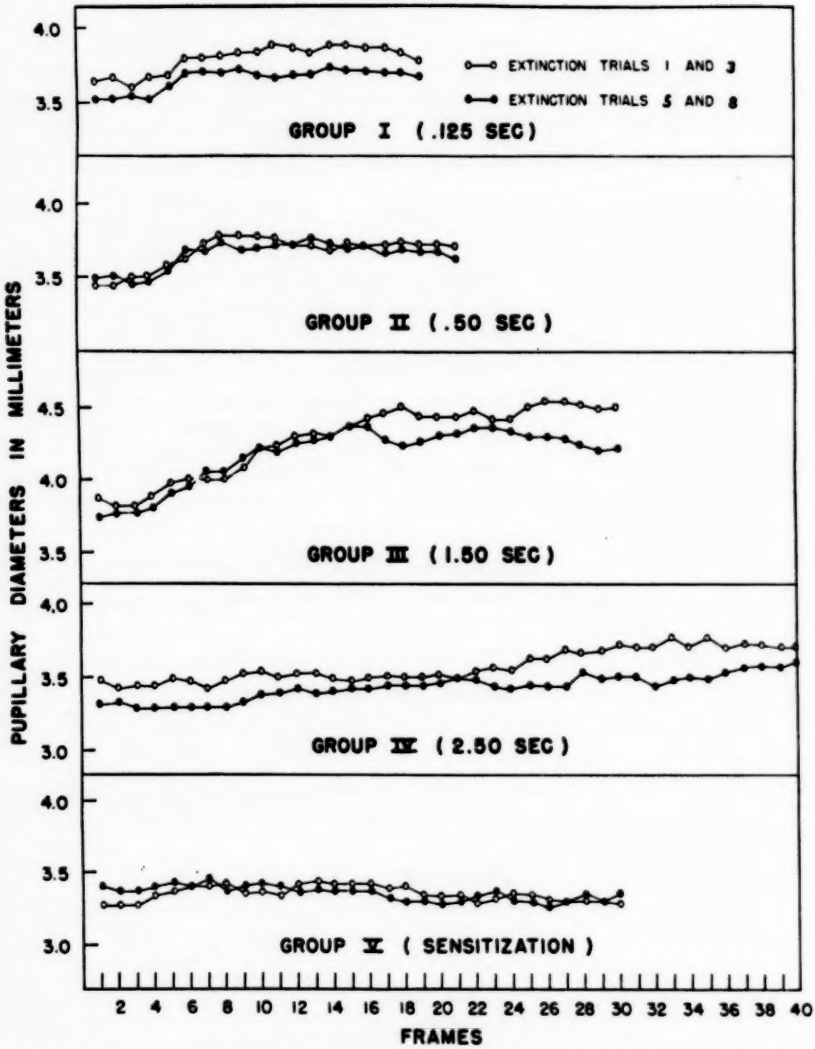

FIG. 2. The mean diameter in millimeters of the pupil during the photographed extinction trials for each of five groups. The first frame coincided with the onset of the CS and 1/10 sec. elapsed between each succeeding frame.

three training phases so that 12 functions were obtained for every S of each group. Before presenting the statistical analyses of these data, it might be useful to illustrate them to some extent. To do this succinctly, a single average function was computed from all the adaptation trials of each group. Also for both test and extinction phases, the first two

and second two photographed trials were averaged for each group. These average curves are presented in Fig. 1 and 2. An inspection of the mean adaptation curves in Fig. 1 reveals that no change in pupil diameter during the CS period occurred in Groups III, IV, and V. A slight increase in pupil size was evoked by the tone in Groups I and II.

The greatest change in pupil size in comparison to the adaptation level occurs in Group III. There appears to be an increase in average pupil diameter in Group II over its adaptation level. It is difficult to make any judgments from these graphs about the performance of Groups I and IV. Group V does not show any change in reactivity of the pupil as a consequence of the sensitization procedure. In Fig 2, the average extinction curves are depicted for each group. Again, the largest change in pupil diameter is manifested by Group III and the second largest change by Group II.

For the purpose of statistical evaluation, a difference score was obtained by subtracting the average diameter of the first three pupils recorded after the onset of the CS from the average diameter of the three largest pupils in succession photographed during any part of the CS period. A difference score of zero would indicate no change in pupil diameter evoked by the CS, and a score greater than zero would occur if the pupil dilated during the CS period. Table 1 presents a summary of the analyses of variance performed on these scores obtained from the adaptation, test, and extinction trials. The only significant effect was due to the differences among the groups during the test and extinction trials.

If a change in pupillary diameter is to be classified as a CR, it must be greater than the change in the pupil evoked by the CS during the adaptation series. A measure of the comparative size of the pupil before and after conditioning was obtained by computing the mean difference between the average dilation response (obtained as described above) of the last two adaptation trials of *S* from his dilation response recorded for each test trial. The mean of these scores obtained from the four test trials for each of the five groups is shown in Fig 3. An analysis of variance of these scores showed that

TABLE 1

ANALYSES OF VARIANCE PERFORMED ON DIFFERENCE SCORES COMPUTED FOR ADAPTATION, CONDITIONING, AND EXTINCTION PHASES

Source	*df*	Adaptation		Conditioning		Extinction	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between groups	24						
Groups	4	.52	1.02	6.64	13.7**	6.46	8.6**
Error (b)	20	.51		.47		.75	
Within groups	75						
Trials	3	.03		.33		.40	2.5
G × T	12	.42		.08		.16	
Error (w)	60	.27		.27		.16	
Total	99						

** $P < .01$.

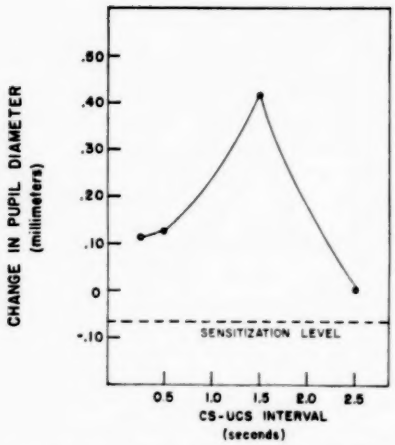

FIG 3. The mean change in diameter of the pupillary response for each of five groups. The dilation response of four of the five groups is plotted above their particular CS-UCS interval indicated on the abscissa. The sensitization level was obtained from the fifth group which did not have the CS and UCS paired during the conditioning series.

the variation due to groups was statistically significant. The null hypothesis was rejected if $P < .05$ for all statistical tests in this study. Individual t tests were performed for all comparisons of the five groups. As might be expected from the previous figures and analyses, Group III showed the largest conditioned dilation response and was statistically different from the other groups. The amplitudes of CR's of the other four groups were not statistically different from each other or from zero.

DISCUSSION

The purpose of this study was to obtain information that would be useful in designing experiments involving the pupillary dilation response. The results partially fulfill this goal. Conditioning was found to occur with a CS-UCS interval of 1.5 sec. An interval of 1.5 sec. has been used in two other experiments in this laboratory (**1, 2**), and in both instances conditioning occurred at a clearly discernible level. Although the statistical analyses showed that the .125- and .50-sec. groups did not manifest a significant dilation response to the CS beyond that evoked before starting the pairing of the CS and UCS, we are reluctant to state that no conditioning occurred at all in these groups. There is an indication that the CS evoked a response in some *S*s in these groups that was greater than the preconditioning magnitude. To some extent, the conditioning curves presented in Fig. 1 and 2 reflect this tendency for greater dilation after initiation of the conditioning procedure. Also, there is a trend toward decreased responsiveness in the pupil to the tone as a function of number of experimental trials in Groups IV and V that was not present to the same degree in Groups I and II. Thus, it is suggested that conditioning of the pupillary dilation response can occur to .50-sec. and perhaps to .125-sec. intervals. There was no evidence for conditioning with a 2.5-sec. interval.

With reference to the question of the value of an optimal CS-UCS interval for the classical conditioning of the dilation response, the results are suggestive rather than definitive. It is possible that the interval found best in this study is a function of the particular experimental procedure followed. Number of conditioning trials or duration of the intertrial interval, or intensity of shock, etc., might influence the efficacy of a particular CS-UCS interval. The data obtained here can only serve as an estimate of the optimal CS-UCS interval for the conditioning of the pupillary dilation response until results of other experiments are reported. As a first estimate, the results are not in agreement with an S-S contiguity theory that would predict an optimal CS-UCS interval of approximately .5 sec. for the classical conditioning of all responses. It appears that the rate or level of conditioning might not be related solely to the temporal separation between the onsets of the CS and UCS, but also to characteristics of the response, such as its latency.

If the level of conditioning is related

to the latency of the UCR, then the reported similarity of the optimal CS-UCS interval for the GSR and skeletal muscle responses must be explained. A hypothesis suggested by J. S. Brown has been presented by Moeller (**9**) to account for this finding. The main premise of the hypothesis is that skeletal muscle responses are conditioned in the relevant GSR situations. It is known that skeletal muscle responses such as coughing, deep breathing, tensing, etc., are followed by a change in magnitude of the GSR. The CS, it is suggested, acquires the capacity as a result of the conditioning procedure to evoke a skeletal muscle response which in turn produces the recorded change in GSR level. If the magnitude of the observed GSR is a function of the level of conditioning of a skeletal muscle response in the experiments reported by White and Schlosberg and by Moeller, then the optimal CS-UCS interval should be approximately .5 sec.

The question arises whether degree of conditioning of the pupillary dilation response might also be determined completely or in part by skeletal muscle activity. Several experiments have been reported in which the dilation response has been conditioned in cats with their skeletal muscle system paralyzed by curare (**1, 3, 5**). It appears unlikely, therefore, that the conditioning of this response is determined completely by the action of skeletal muscles, especially of the type discussed by Moeller. However, no statement can be made at this time as to whether degree of conditioning of this response to different CS-UCS intervals is completely independent of skeletal muscle activity. To provide information bearing on this problem, it would be necessary to compare the level of pupillary response conditioning obtained with different CS-UCS intervals in normal and curarized subjects. If the GSR were also measured in such an experiment, data pertaining to the hypothesis suggested by J. S. Brown could also be obtained.

Summary

Four groups of *S*s were presented with different CS-UCS intervals in an attempt to modify the pupillary dilation response by the classical conditioning method. A fifth group was a control for sensitization effects. The highest level of conditioning was found to be manifested in the group that had a 1.5-sec. interval. There was some indication that conditioning occurred in the .125- and .50-sec. groups. No conditioning was found for the groups with the 2.5-sec. CS-UCS interval. The data were interpreted as not supporting the S-S theory of learning discussed by White and Schlosberg.

References

1. Gerall, A. A., & Obrist, P. A. Conditioning of the pupillary dilation response in normal and curarized cats. Paper read at Eastern Psychol. Ass., New York, April, 1957.
2. Gerall, A. A., Sampson, P. B., & Boslov, G. L. Classical conditioning of the human pupillary dilation response. *J. exp. Psychol.*, 1957, **54,** 467–474.
3. Girden, E. The dissociation of pupillary conditioned reflexes under erythroidine and curare. *J. exp. Psychol.*, 1942, **31,** 322–332.
4. Glazer, V. D. Pupillary dilation reaction. *Physiol. Zh.*, 1952, **38,** 604–610.
5. Harlow, H. F., & Stagner, R. Effect of complete striate muscle paralysis upon the learning process. *J. exp. Psychol.*, 1933, **16,** 283–294.
6. Hilgard, E. R., Dutton, C. E., & Helmick, J. S. Attempted pupillary conditioning at four stimulus intervals. *J. exp. Psychol.*, 1949, **39,** 683–689.
7. Kimble, G. A. Classical conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37,** 1–36.
8. McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45,** 417–422.
9. Moeller, G. The CS-UCS interval in conditioning the GSR. *J. exp. Psychol.*, 1954, **48,** 162–166.
10. White, C. T., & Schlosberg, H. Degree of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43,** 357–362.
11. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.
12. Young, F. A., & Biersdorf, W. R. Pupillary contraction and dilation in light and darkness. *J. comp. physiol. Psychol.*, 1954, **47,** 264–268.

(Received July 2, 1957)

KEEP POSTED

Read

CONTEMPORARY PSYCHOLOGY

A Journal of
Reviews
Criticism
Opinion

Can you keep posted on the ever-increasing number of books now published in your field—books on social and abnormal psychology, clinical psychology, consulting psychology, psychotherapy, personality, mental health?

Let *CP* help by bringing you each month selective reviews by competent reviewers of important books, controversial books, worthwhile books, useful books.

Read the Editor's page, CP SPEAKS, for news of the publishing world, comment on communication in general and in particular, discussion of reviewing and of criticism.

Subscription, $8.00
(Foreign, $8.50)

Single copy,
$1.00

Send subscription orders to:

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

Publication Manual
of the American Psychological Association
1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. All scientists who are writing for publication will find the Publication Manual an indispensable guide.

Price $1.00

Discounts on quantity orders over 50 copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION
Publications Office, 1333 Sixteenth Street, N.W., Washington 6, D. C.

AMERICA'S PSYCHOLOGISTS

A Survey of a Growing Profession

KENNETH E. CLARK

Associate Professor of Psychology

University of Minnesota

A report on a three-year study of American Psychologists, Sponsored by the Policy and Planning Board of the American Psychological Association and the National Science Foundation. The volume presents a view of a growing profession in the mid-twentieth century by describing the people active in psychology. $1.00

Order from

American Psychological Association, Inc.

1333 Sixteenth Street, N.W.

Washington 6, D.C.